

Funcionamiento diferencial del ítem en la evaluación de la sintomatología TDAH en función del género y el formato de calificación

Differential Item Functioning in the Assessment of ADHD Symptoms Based on Gender and Rating Format

Recibido: diciembre 31 de 2011 | Revisado: agosto 18 de 2012 | Aceptado: diciembre 26 de 2012

BENITO ARIAS MARTÍNEZ*

Universidad de Valladolid, España

VÍCTOR B. ARIAS GONZÁLEZ**

Universidad de Talca, Chile

LAURA ELÍSABET GÓMEZ SÁNCHEZ***

Universidad de Oviedo, España

MARÍA ANGÉLICA INMACULADA CALLEJA****

Universidad de Valladolid, España

RESUMEN

El objetivo de este estudio se centró en poner a prueba la invarianza de la sintomatología del Trastorno por Déficit de Atención con Hiperactividad (TDAH) en función del género, en una muestra de 634 niños. Se comprobó, en primer lugar, el ajuste de cinco modelos factoriales mediante análisis factorial confirmatorio, y se utilizó la regresión logística ordinal como método de estimación del funcionamiento diferencial del ítem (DIF), tanto uniforme como no uniforme. Los resultados pusieron de manifiesto que: (a) el modelo que presentó mejor ajuste fue el de tres factores correlacionados y (b) no existe DIF en función del género de los niños evaluados, ni en la modalidad de calificación ordinal (escala de 1 a 4) ni en la modalidad de calificación binaria (0-1) de los ítems. Estos resultados refrendan el hecho de que en el *Diagnostic and Statistical Manual of Mental Disorders, DSM-IV-TR*, no se establezcan criterios diferenciales para el diagnóstico del TDAH en niños y niñas.

Palabras clave autores

Funcionamiento diferencial del ítem, DIF, instrumentos de evaluación, TDAH, regresión logística ordinal.

Palabras clave descriptores

Psicometría, Psicología cuantitativa, Psicología clínica.

ABSTRACT

The aim of this paper was to test the invariance of Attention Deficit Hyperactivity Disorder (ADHD) symptoms in a sample of 634 children attending to gender. We firstly examined the fit to the data of five models using confirmatory factor analysis, and ordinal logistic regression was used as a method of estimation of both uniform and non-uniform differential item functioning (DIF). The results showed that (a) the three correlated factors model was the best fit model, and (b) there was no gender-specific DIF, either in the ordinal rating form (scale of 1 to 4) or the binary rating form (0-1). These results support that the *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR* does not establish ADHD differential diagnosis criteria for boys and girls.

Key words authors

Differential item functioning, DIF, assessment instruments, ADHD, ordinal logistic regression.

Key words plus

Psychometry, Quantitative Psychology, Clinical Psychology.

doi:10.11144/Javeriana.UPSY12-3.fdie

Para citar este artículo: Arias, B., Arias, V. B., Gómez, L. E. & Calleja, M. A. I. (2013). Funcionamiento diferencial del ítem en la evaluación de la sintomatología TDAH en función del género y el formato de calificación. *Universitas Psychologica*, 12(3), 779-796. doi:10.11144/Javeriana.UPSY12-3.fdie

* Universidad de Valladolid, España. Profesor Titular. Departamento de Psicología. ResearcherID: D-7925-2013. E-mail: barias@psi.uva.es

** Universidad de Talca, Chile. Profesor Asistente. Facultad de Psicología. ResearcherID: H-6294-2013. E-mail: viarias@utalca.cl

*** Universidad de Oviedo, España. Profesora Ayudante Doctora. Departamento de Psicología. ResearcherID: B-5013-2011. E-mail: gomezlaura@uniovi.es

**** Universidad de Valladolid, España. Profesora Titular. Departamento de Psicología. ResearcherID: H-6425-2013. E-mail: inca@psi.uva.es

El Trastorno por Déficit de Atención con Hiperactividad (TDAH) es, según el sistema clasificatorio *Diagnostic and Statistical Manual of Mental Disorders - DSM-IV-TR* (American Psychiatric Association, 2002), uno de los más frecuentes entre los trastornos del comportamiento infantil. Su prevalencia se cifra, de acuerdo con el citado manual, entre el 3 y el 7 %; si bien estas cifras varían notablemente en función del uso de diferentes criterios diagnósticos, del tipo de instrumento de evaluación que se utilice (entrevista, cuestionarios, escalas de clasificación, observación directa), del tipo de muestra (clínica, subclínica o comunitaria), de quiénes sean los informantes (padres, maestros) y de otras variables, tales como las condiciones geográficas, étnicas, culturales o sociales (Amador, Forns & Martorell, 2001; Barkley & Murphy, 2006; Bauermeister, Berrios, Jimenez, Acevedo & Gordon, 1990; Cardo & Servera, 2005; Reid et al., 1998). En general, el rango de prevalencia en los diversos estudios epidemiológicos viene a coincidir con el *DSM-IV-TR* (Benjumea & Mojarro, 1993; Cardo & Servera, 2005; Cardo, Servera & Llobera, 2007; Polanczyk, De Dima, Horta, Biederman & Rohde, 2007; Rodríguez et al., 2009; Spencer, Biederman & Mick, 2007).

En cuanto a los datos epidemiológicos en función del género, se estima que el trastorno presenta una prevalencia nueve veces mayor en niños que en niñas en muestras clínicas, y tres veces mayor en muestras comunitarias (APA, 2002; Arnold, 1996; Barkley, 2006). Si tenemos en cuenta los subtipos, la ratio es de 2:1 en el desatento y de 4:1 en el hiperactivo-impulsivo, según la *American Academy of Child and Adolescent Psychiatry* (1997), siempre constituyendo los varones el grupo más numeroso.

Diversos estudios han señalado que los síntomas asociados al TDAH varían a lo largo del desarrollo en frecuencia e intensidad, según la edad y el género (Amador, Forns & Martorell, 2001; Amador, Forns, Guàrdia & Però, 2006a, 2006b; DuPaul et al., 1997; DuPaul et al., 1998; Hinshaw, Han, Erhardt & Huber, 1992). Así, los síntomas de hiperactividad-impulsividad son más frecuentes en la edad infantil, en comparación con los síntomas de desatención, tanto cuando los informantes son los padres como cuando son los maestros (Amador

et al., 2006a; DuPaul et al., 1997; DuPaul et al., 1998) y también son más frecuentes en estos niños que en los de mayor edad (Amador et al., 2006a). Los síntomas de desatención son más frecuentes en los niños de 5 a 12 años y menos frecuentes en los de 4 a 5 y 14 a 18 años (DuPaul et al., 1998). Por otra parte, los niños presentan mayor número de síntomas de TDAH que las niñas (Amador et al., 2006a; DuPaul et al., 1998).

Pese a las diferencias en función del género mencionadas, el *DSM-IV-TR* (APA, 2002) utiliza idénticos criterios para el diagnóstico en niños y niñas, estableciendo un único punto de corte para todos ellos (i. e., seis o más síntomas durante un mínimo de seis meses y en al menos dos ambientes distintos en cualquiera de las dos dimensiones de desatención o hiperactividad-impulsividad). Parece claro que se deben realizar estudios que determinen si es o no necesario establecer puntos de corte diferentes en cuanto a la sintomatología en función del género, incluyendo a todo el espectro de edades, desde la primera infancia a la adolescencia, tanto en muestras comunitarias como subclínicas y clínicas.

Carecemos, por otra parte, de investigaciones que hayan llegado a conclusiones suficientemente sólidas respecto a si los niños y las niñas responden de manera diferencial en función del contexto social y del contexto del desarrollo, por lo que se requiere conocer esas manifestaciones a lo largo de todo el período evolutivo y sus variaciones con respecto a los datos normativos.

Antes de los 5 o 6 años es difícil el diagnóstico, debido a que la inquietud, la impulsividad o la desatención pueden ser normales desde el punto de vista evolutivo, y a que las exigencias del contexto escolar no suelen ser muy elevadas. Si a esto le añadimos que las pruebas neurocognitivas suelen estar diseñadas para niños de mayor edad, y que no hay una perspectiva suficiente para comprobar si sus comportamientos desaparecen con la intervención educativa ordinaria de los colegios, el diagnóstico puede resultar extremadamente difícil, cuando no erróneo. Sin embargo, sería conveniente incluir en los estudios de investigación niños y niñas de menor edad, pues facilitaría la labor de un buen diagnósti-

co de marcadores precoces del trastorno, y la puesta en marcha de programas de intervención temprana.

Se han desarrollado bastantes instrumentos para el diagnóstico del TDAH, basados en la sintomatología *DSM-IV-TR*, como la ADHD Rating Scale-IV (DuPaul et al., 1998), la DSM-IV ADHD Rating Scale - DARS (Burns, Walsh & Gomez, 2003; Gomez, 2007; Gomez, Harvey, Quick, Scharrer & Harris, 1999); la Disruptive Behavior Rating Scale (Barkley, 2006), la Vanderbilt ADHD Parent Ratings Scale - VADPRS (Wolraich et al., 2003), la ADHD Symptom Checklist-4 (Gadow & Sprafkin, 1997) o la SNAP-IV Rating Scale (Swanson, 2010), entre otros. Tales escalas reproducen con más o menos exactitud los 18 enunciados originales, además de agruparlos en las dos subescalas de déficit de atención e hiperactividad-impulsividad, cada una formada por nueve ítems. Algunas de ellas se presentan en dos formatos (para padres y maestros) con el objeto de poder evaluar la sintomatología en dos contextos diferentes y, por lo general, se solicita a los respondientes que evalúen la frecuencia de los síntomas en los últimos seis meses. El formato de respuesta suele ser una escala de frecuencia de cuatro puntos que va de 0 (*nunca* o *raramente*) a 3 (*siempre* o *muchas veces*). Puntuaciones directas entre 12 y 18 (o superiores a 6 si la escala es binaria) en una o ambas subescalas denotarían la presencia del trastorno.

Por otro lado, los resultados de los estudios centrados en el análisis de la estructura factorial suelen coincidir en la existencia de dos factores correlacionados, a la par que se pone de manifiesto una mayor diferenciación entre la fuente de información (padres, maestros) que entre las propias variables latentes (desatención, hiperactividad-impulsividad) (Amador, Forns, Guàrdia & Peró, 2005; Balluerka, Gómez, Stock & Caterino, 2000; Cardo et al., 2007; Gomez, Burns, Walsh & Hafetz, 2005; Servera & Cardo, 2007).

Puesto que, como más arriba se mencionó, en el manual *DSM-IV-TR* (APA, 2002) se asignan idénticos criterios para diagnosticar el trastorno en niños y niñas, se deduce que los síntomas son invariantes entre ambos géneros. A partir de tal deducción, el presente estudio se ha planteado

como objetivo poner a prueba dicha invarianza en la sintomatología de niños y niñas, utilizando dos formatos de respuesta (ordinal y binario). Se han analizado ambos formatos debido a que el primero es usualmente utilizado en contextos educativos y de investigación, en tanto que el segundo se utiliza frecuentemente en entrevistas diagnósticas. La invarianza de género se entiende como la equivalencia en las puntuaciones de niños y niñas cuando los sujetos evaluados han sido igualados en la variable latente (i. e., TDAH en sus dos modalidades de desatención e hiperactividad-impulsividad).

Método

Participantes

Participaron en este estudio 634 niños (314 varones y 320 mujeres) con una media de edad de 66.59 meses ($DE = 5.92$), distribuidos en un rango de 56 a 83 meses. La media de edad de los niños fue de 66.69 meses ($DE = 5.92$) y la de las niñas de 66.49 ($DE = 5.93$). En la Figura 1 puede consultarse la distribución de la muestra por género y edad. No se apreciaron diferencias significativas en la variable edad ($t(632) = 0.419$; $p = 0.675$), atendiendo al género de los participantes. Tampoco las frecuencias observadas en función del género fueron significativamente distintas a las esperadas bajo el supuesto de equiprobabilidad ($\chi^2(1) = 0.057$; $p = 0.812$) en el conjunto de la muestra. Los niños asistían a 10 centros escolares de la Comunidad de Castilla y León y fueron seleccionados de forma incidental. En todos los casos se contó con el consentimiento informado por parte de los padres o representantes legales.

Instrumento

Se convirtieron los 18 síntomas propuestos por el *DSM-IV-TR* (APA, 2002) en otros tantos ítems y se incluyó una escala de respuesta de frecuencia de 4 puntos (1 = *casi nunca* o *muy pocas veces*, 2 = *algunas veces*, 3 = *bastantes veces* y 4 = *muchas veces* o *casi siempre*). La formulación de los ítems coincidió de forma literal con la versión española

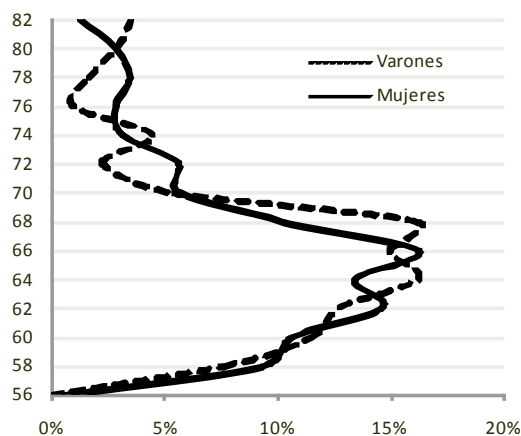


Figura 1. Distribución de la muestra por género y edad.
Fuente: elaboración propia.

del DSM-IV-TR, con la salvedad de que se suprimió la expresión adverbial ‘a menudo’ con que comienza la redacción de todos los síntomas, dado que el uso de tal expresión anularía *de facto* la escala de frecuencia. Tomando como base los resultados del análisis factorial confirmatorio (véase el apartado de Resultados), se consideraron para los ulteriores análisis tres grupos de ítems: el primero del 1 al 9 (síntomatología relativa a la desatención), el segundo del 10 al 15 (síntomatología relativa a la hiperactividad) y el tercero del 16 al 18 (síntomatología relativa a la impulsividad). Como más arriba se ha indicado, es común, a la hora de intentar evaluar la sintomatología del TDAH que propone el DSM-IV-TR, trasladar tales síntomas a escalas similares a la aquí utilizada (Amador et al., 2005; Barkley, 2006; Barkley & Murphy, 2006; Burns et al., 2003; Dupaul et al., 1998; Gomez, 2007; Gadow & Sprafkin, 1997; Gomez et al., 1999; Swanson, 2010; Wolraich et al., 2003).

Procedimiento

Una vez confeccionada la escala como se ha explicado en el apartado anterior, se trasladó a formato electrónico mediante el programa LimeSurvey v. 1.87 (LimeSurvey, 2009) con el fin de que se pudiera cumplimentar por internet. Se estima que esta modalidad tiene varias ventajas sobre el método tradicional de ‘lápiz y papel’, entre otras, que el

instrumento puede cumplimentarse en cualquier momento, que las respuestas se almacenan automáticamente en la base de datos del servidor, y que se generan dos archivos –uno en formato .DAT y otro en formato de sintaxis .SPS– que simplifican considerablemente la tabulación de los datos para su posterior análisis con cualquier programa informático. Se hizo contacto a continuación con los profesores de los centros escolares y se les envió la información necesaria sobre la investigación, junto con un modelo de consentimiento informado que debían aceptar los padres o representantes legales de los alumnos. Se dio un plazo de dos semanas para cumplimentar las escalas, pasado el cual se recuperó la información de la base de datos y se procedió a su análisis. Se desearon 29 cuestionarios por estar incompletos. Con los datos válidos ($N = 634$) se construyeron dos ficheros: en el primero se conservaron las puntuaciones originales y en el segundo se transformaron a binarias (1, 2 = 0; 3, 4 = 1) con la finalidad de poder realizar una comprobación posterior sobre el comportamiento de los ítems en ambos tipos de formato.

Análisis de datos

Se llevaron a cabo en este estudio tres clases de análisis. En primer lugar, para comprobar la hipótesis de la estructura factorial de la escala, se realizó un análisis factorial confirmatorio mediante el cual se comparó el ajuste de la estructura de cinco modelos uni y multifactoriales, tanto de primer como de segundo orden. En segundo lugar, se han comparado las tres matrices de correlaciones (policóricas entre los ítems ordinales, tetracóricas entre los binarios y de Pearson entre las puntuaciones totales en los factores). Por último, se realizó un análisis de regresión logística ordinal para poner a prueba el posible funcionamiento diferencial de los ítems.

Junto con el estadístico de Mantel-Haenszel, la regresión logística se ha venido confirmando como una de las técnicas más eficaces para detectar el DIF (Clauser & Mazor, 1998; Ferreres, Fidalgo & Muñoz, 2000; Gelin, Carleton, Smith & Zumbo, 2004; Hidalgo & Gómez, 2000, 2003; Hidalgo, Gómez & Padilla, 2005; Hidalgo & López, 2004; Hosmer

& Lemeshow, 1989; Jodoin & Gierl, 2001; Swaminathan & Rogers, 1990; Whitmore & Schumaker, 1999) en ítems tanto dicotómicos como politómicos, especialmente cuando las muestras tienen un tamaño suficientemente grande.

Los análisis DIF sobre las puntuaciones tanto ordinales como binarias se llevaron a cabo utilizando el método de regresión logística ordinal (RLO) propuesto por Zumbo (1999), junto con el cálculo del correspondiente tamaño del efecto (Jodoin & Gierl, 2001). La regresión logística se basa en un modelo estadístico de la probabilidad de responder correctamente a un ítem por pertenencia a un grupo (varones o mujeres, en nuestro caso) y un criterio o variable condicional, identificada normalmente con la puntuación total obtenida en la escala analizada (en el presente estudio, las obtenidas en las subescalas de Desatención, Hiperactividad e Impulsividad). El procedimiento utiliza como variable dependiente la respuesta a cada uno de los ítems y, como variables independientes (a) el grupo de pertenencia, (b) la puntuación total en la escala y (c) la interacción entre el grupo de pertenencia y la puntuación total.

El método proporciona una prueba del DIF condicional a la relación entre la respuesta a cada ítem y la puntuación total, poniendo a prueba los efectos del grupo (DIF uniforme) y la interacción grupo-total (DIF no uniforme). En consecuencia, la ecuación de la regresión logística ordinal para ítems binarios es:

$$Y = b_0 + b_1\theta + b_2G + b_3(\theta * G) \quad (1)$$

donde Y es el logaritmo natural de la *odds ratio*, θ es la puntuación total en la variable latente evaluada, b_0 la intercepción, b_1 el coeficiente para la puntuación observada, b_2 el coeficiente para el grupo de pertenencia, b_3 la interacción entre puntuación y grupo y G el grupo de pertenencia. En el caso de ítems ordinales, la ecuación es:

$$y^* = b_0 + b_1\theta + b_2G + b_3(\theta * G) + \varepsilon_i \quad (2)$$

donde y^* es una variable latente continua con distribución aleatoria que gobierna las respuestas a los ítems y les proporciona una métrica (en esencia, podría definirse como la propensión o 'disponibilidad' no observada que un individuo tiene para dar

una respuesta a un ítem), y la distribución de los χ_i tiene una media de 0 y una varianza de: $\pi^2/3$

De acuerdo con Zumbo (1999), el análisis RLO requiere un mínimo de 200 sujetos, requisito que cumple sobradamente la muestra utilizada (314 varones y 320 mujeres). El método implica el cálculo del DIF para cada ítem por separado en cada uno de los dos procedimientos de puntuación (ordinal y binario) utilizados.

Las variables independientes se introducen en la ecuación siguiendo la jerarquía que se especifica en los párrafos siguientes. El orden de entrada es relevante, toda vez que permite al investigador determinar el efecto principal del género y la variación adicional explicada por la interacción GRUPO x TOTAL.

1. En el paso 1 se introduce la variable condicional (i. e., puntuación total en la subescala correspondiente –Desatención, Hiperactividad o Impulsividad–).
2. En el paso 2, la variable de agrupación (el género de los niños evaluados en nuestro caso, codificado como 1 para los varones y 2 para las mujeres). En este paso se evalúa el efecto de la variable de agrupación manteniendo constante el efecto de la variable condicional.
3. Finalmente, en el paso 3, el término de interacción (i. e., grupo de pertenencia x puntuación total en cada subescala). Este paso describe si la diferencia entre varones y mujeres en su calificación en el ítem varía a través del continuo de la variable latente (i. e., niveles bajos a niveles altos en Desatención, Hiperactividad o Impulsividad).

Cada uno de los pasos proporciona el valor de χ^2 junto con el correspondiente valor R^2 asociado (que se usa como prueba estadística de la existencia de DIF).

Se considera que un ítem muestra DIF entre los grupos (género) cuando se cumplen dos condiciones (Gelin et al., 2004; Gelin & Zumbo, 2003): en primer lugar, el valor de $\chi^2_{(3)}$ en el paso 3 deberá ser significativamente superior al valor de $\chi^2_{(1)}$ en el paso 1; en otras palabras, el valor de $\Delta\chi^2_{(2)}$ deberá ser superior a 9.21 (que corres-

ponde a $p = 0.01$). En segundo lugar, el tamaño del efecto de dicha diferencia entre los valores de χ^2 deberá tener una magnitud suficiente. El uso de un nivel de confianza del 99 % obedece a que cuando se dispone de muestras grandes, diferencias pequeñas pueden producir diferencias significativas en χ^2 . Zumbo (1999) ha sugerido que con métodos de calificación ordinales y binarios pueden usarse las diferencias en el valor de R^2 en el paso 3 sobre el paso 1 para evaluar el tamaño del efecto de la diferencia de los valores correspondientes de χ^2 . Para considerarse significativa, la diferencia entre los valores de R^2 (denominada ‘tamaño del efecto de Zumbo-Thomas’) debería superar el valor de 0.13.

En el caso de que se encuentren ítems con DIF, pueden examinarse los resultados de la RLO para asegurarse de si el DIF es uniforme o no uniforme. El DIF uniforme tiene lugar cuando uno de los grupos obtiene puntuaciones consistentemente más elevadas que el otro. En otras palabras, la relación estadística entre la respuesta al ítem y el grupo de pertenencia es constante para todos los niveles de la variable latente y el efecto de la pertenencia al grupo –G– es estadísticamente significativo, en tanto que la interacción θG no ejerce efecto sobre el ítem.

El DIF no uniforme implica que la interacción θG es estadísticamente significativa y denota que en determinados niveles de la variable latente subyacente, un grupo tiene puntuaciones más altas, mientras que en otros niveles de la variable latente

sucede lo contrario. En términos de teoría de respuesta a los ítems, el DIF uniforme significa que las curvas características del ítem (que muestran la relación entre la puntuación en el ítem y la puntuación en la variable latente subyacente) de los dos grupos no se cruzan en ningún nivel del continuo que representa la variable latente, mientras que el DIF no uniforme significa que las curvas características de los ítems de ambos grupos se cruzan en algún punto.

Cuando se usa el método de la regresión logística ordinal para calcular el DIF, el modelo completo (Full Model) proporciona tanto el DIF uniforme como el no uniforme, es decir, la diferencia de χ^2 entre los pasos 3 y 1 proporciona una prueba simultánea de DIF uniforme y no uniforme. Por definición, el DIF uniforme es un efecto principal significativo para los grupos, en tanto que el DIF no uniforme es un efecto significativo de la interacción grupo x puntuación total. De este modo, diferencias sustanciales entre los valores R^2 entre los pasos 1 y 2 pueden usarse como indicadores de DIF uniforme. Este valor indica la variación única atribuible a las diferencias entre los grupos por encima de la variable condicional. Cuando se encuentra DIF uniforme, puede utilizarse la *odds ratio* para determinar la dirección del DIF, esto es, qué grupo tiene más probabilidad de obtener una puntuación alta en un ítem. Una diferencia sustancial de los valores R^2 entre los pasos 2 y 3 indicaría la presencia de DIF no uniforme.

TABLA 1
Índices de ajuste de los modelos

	M1	M2	M3	M4	M5
χ^2	1293.422	504.028	425.564	503.964	401.917
gl	135	134	132	133	132
p	<0.001	<0.001	<0.001	<0.001	<0.001
CFI	0.947	0.983	0.987	0.983	0.988
TLI	0.94	0.981	0.985	0.981	0.986
RMSEA	0.116	0.066	0.059	0.066	0.056
(IC 90%)	0.111 ; 0.122	0.06 ; 0.072	0.053 ; 0.065	0.07 ; 0.072	0.05 ; 0.063
SRMR	0.0905	0.068	0.059	0.068	0.066

Nota. χ^2 = Ji cuadrado Escalado de Satorra-Bentler; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.
Fuente: elaboración propia.

FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM EN LA EVALUACIÓN DE TDAH
EN FUNCIÓN DEL GÉNERO Y EL FORMATO DE CALIFICACIÓN

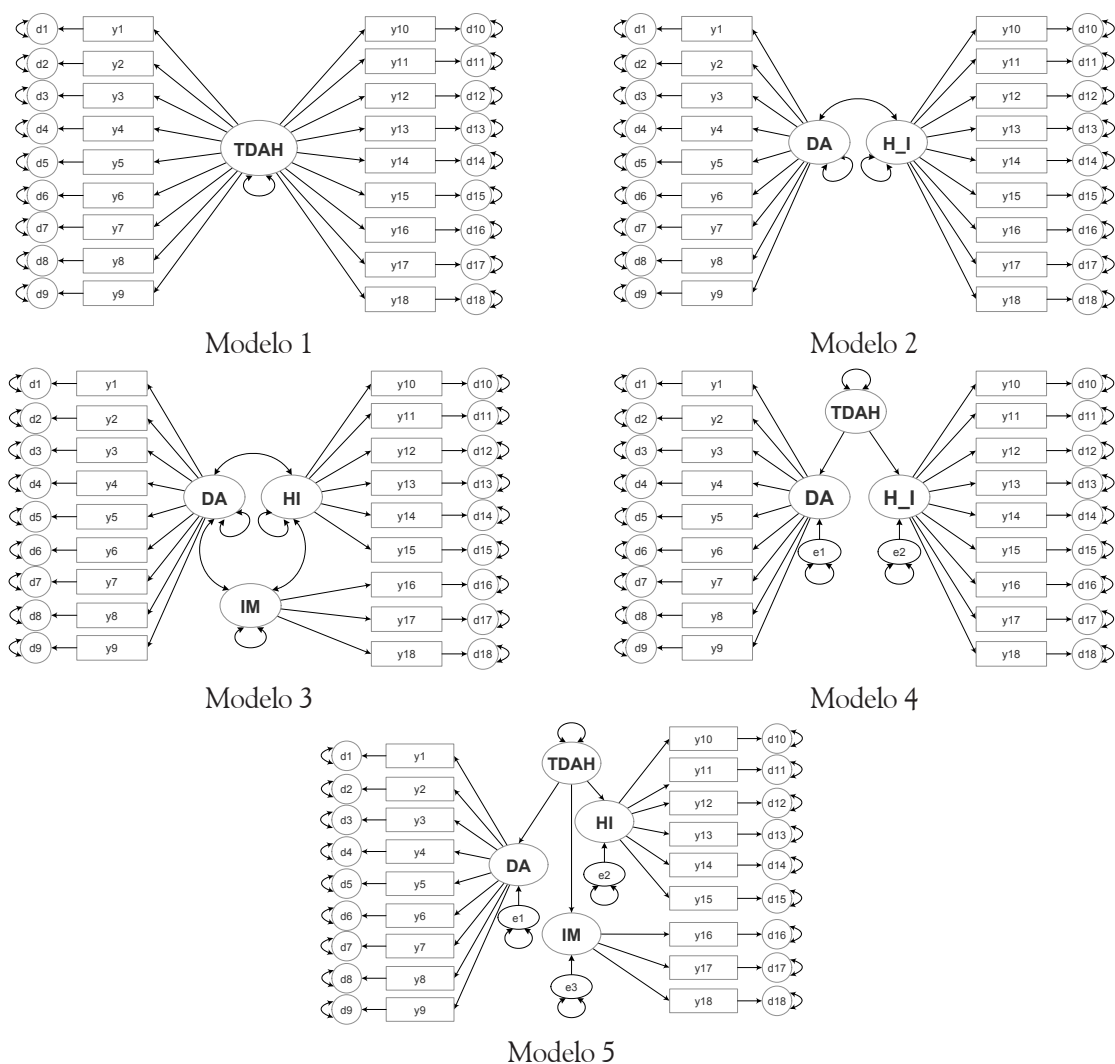


Figura 2. Esquema de los modelos puestos a prueba.

TDAH = Trastorno por Déficit de Atención con Hiperactividad; DA = Déficit de Atención; HI = Hiperactividad; IM = Impulsividad; H_I = Hiperactividad/ Impulsividad.

Fuente: elaboración propia.

Resultados

Análisis factorial confirmatorio

Se sometieron a prueba los cinco modelos que vienen esquematizados en la Figura 2: Modelo 1 (unifactorial); Modelo 2 (dos factores correlacionados); Modelo 3 (tres factores correlacionados); Modelo 4 (modelo jerárquico con dos factores de primer orden y uno de segundo orden) y Modelo 5 (tres factores de primer orden y uno de segundo orden). El análisis se llevó a cabo utilizando el programa

LISREL v. 8.8 (Scientific Software International, 2008) sobre la matriz de varianzas-covarianzas asintóticas entre las 18 variables observadas. Dada la estructura ordinal de las escalas y la ausencia de normalidad multivariada ($\chi^2 = 1421.36; p < 0.001$), se utilizó DWLS como método de estimación de los parámetros. Los resultados de los análisis se sintetizan en la Tabla 1.

Como puede apreciarse en la Tabla 1, el modelo que mostró mejor ajuste fue el de tres factores de primer orden y uno de segundo orden (Modelo 5), seguido del de tres factores correlacionados (Mo-

delo 3). En consecuencia, se optó por realizar el análisis del DIF sobre la solución trifactorial.

Comprobación de la unidimensionalidad de las subescalas

Una de las asunciones del análisis RLO es que los ítems que componen una subescala tengan una estructura unidimensional. Utilizando el programa Mplus v. 6.12 (Muthén & Muthén, 2011) para calcular las correlaciones policóricas, se analizó la unidimensionalidad de las tres subescalas mediante el cálculo de los coeficientes ordinales de consistencia interna (α de Cronbach) y θ (Elosua & Zumbo, 2008; Zumbo, Gadermann & Zeisser, 2007), la media de las correlaciones ítem-total y los valores propios de las soluciones factoriales sin rotar mediante el método de componentes principales para las dos modalidades de calificación (ordinal y binaria). Los resultados que se muestran en la Tabla 2 indican que: (a) el valor α de Cronbach para datos ordinales en todos los casos es elevado (superior a 0.9) al igual que el valor θ ordinal (en torno a 0.9), así como las correlaciones ítem-total (superiores a 0.5, con la excepción de la escala de Desatención en la modalidad de calificación binaria).

El coeficiente α ordinal se ha calculado mediante la fórmula:

$$\alpha = (p / p - 1) \left[\left(p(\bar{f})^2 - \bar{f}^2 \right) / \left(p(\bar{f})^2 + \bar{u}^2 \right) \right] \quad (3)$$

donde \bar{f} es la media de las p saturaciones factoriales (i. e., número de ítems); \bar{f}^2 es la media de los cuadrados de las p saturaciones factoriales y \bar{u}^2 es la media de las p unicidades.

El valor θ ordinal se ha obtenido mediante la fórmula:

$$\Theta = \left[(p / p - 1) \right] \left[1 - (1 / \lambda_1) \right] \quad (4)$$

donde p es el número de saturaciones factoriales y λ_1 es el valor propio mayor, resultante del análisis factorial de la matriz de correlaciones policóricas entre los ítems. Puesto que la fiabilidad de las esca-

las multidimensionales tiende a ser infraestimada por el coeficiente α , y el grado de infraestimación es mayor cuanto más elevadas son las correlaciones entre las subescalas, se calculó el α estratificado (Abell, Springer & Kamata, 2009; Cronbach, Schönemann & McKie, 1965) de acuerdo con la ecuación:

$$\alpha_s = \sum_i \sigma_{o_i}^2 (1 - \alpha_i) / \sigma_x^2 \quad (5)$$

donde $\sigma_{o_i}^2$ es la varianza de la puntuación total en la subescala i , α_i es el coeficiente α de cada subescala y σ_x^2 es la varianza de la escala total.

Otro argumento que redundaba en la unidimensionalidad de las tres subescalas estriba en que los valores propios del primer factor son notoriamente superiores a los del segundo factor, y así queda reflejado en las razones correspondientes. Estas alcanzan valores superiores en la subescala de Desatención, observándose los valores más bajos en la subescala de Impulsividad. El cálculo de la fiabilidad de los ítems, considerada como la correlación múltiple al cuadrado obtenida mediante la estimación WLS, presentó una media de 0.73 para la escala de Desatención, de 0.78 para la de Hiperactividad y de 0.69 para la de Impulsividad (los valores individuales pueden consultarse en las Tablas 3, 4 y 5).

Comparación de las matrices de correlación

Se compararon las dos matrices de correlaciones policóricas entre los ítems ordinales, tetracóricas entre los binarios y producto-momento de Pearson entre las puntuaciones totales en los factores mediante los dos procedimientos que se señalan a continuación.

El primero consistió en comparar las estructuras de correlación utilizando ecuaciones estructurales (Jöreskog & Sörbom, 2008; Kline, 2010; Von Eye & Mun, 2005). Se trata de poner a prueba la hipótesis de que las dos matrices de correlación son invariantes, i. e., $\theta^{(1)} = \phi^{(2)}$.

Los resultados en el caso de los 18 ítems (Tabla 3) denotan un ajuste satisfactorio del modelo a los datos. Las diferencias en las contribuciones a χ^2

FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM EN LA EVALUACIÓN DE TDAH
EN FUNCIÓN DEL GÉNERO Y EL FORMATO DE CALIFICACIÓN

TABLA 2
Unidimensionalidad de las subescalas

	Puntuación ordinal			Puntuación binaria		
	DA	HI	IM	DA	HI	IM
α de Cronbach ordinal	0.96	0.97	0.9	0.98	0.96	0.92
α estratificado		0.979			0.997	
θ ordinal	0.92	0.89	0.79	0.95	0.92	0.85
Media correlaciones ítem-total	0.546	0.597	0.614	0.446	0.672	0.511
Valores propios Factor 1	5.49	3.9	2.1	6.3	4.3	2.32
Valores propios Factor 2	0.71	0.63	0.54	0.63	0.62	0.42
Razón eigen1/eigen2	7.73	6.19	3.89	10	6.94	5.52

Nota: DA = Déficit de atención; HI = Hiperactividad; IM = Impulsividad.
Fuente: elaboración propia.

TABLA 3
Comparación de las matrices de correlación

	Modalidad Ordinal				Modalidad Binaria			
	18 ítems		3 factores		18 ítems		3 factores	
	Varones	Mujeres	Varones	Mujeres	Varones	Mujeres	Varones	Mujeres
Contribución a χ^2	268.23	228.69	1.06	0.96	150.71	129.74	3.3	2.78
% Contrib. a χ^2	53.98	46.02	52.4	47.6	53.74	46.26	54.33	45.67
SRMR	0.048	0.048	0.012	0.012	0.034	0.06	0.019	0.019
GFI	0.93	0.93	1	1	0.99	0.92	0.99	0.99
GL	171		6		171		6	
χ^2	496.92 ($p < 0.001$)		2.02 ($p = 0.92$)		280.45 ($p < 0.001$)		6.08 ($p = 0.410$)	
RMSEA (IC90%)	0.072 (0.064 ; 0.08)		0 (0 ; 0.026)		0.053 (0.044 ; 0.062)		0.0037 (0 ; 0.0073)	
P-close	0		0.99		0.29		0.79	
TLI	0.97		1		0.99		1	
CFI	0.98		1		0.99		1	

Nota. GFI = Goodness of Fit Index; χ^2 = Minimum Fit Function Chi-Square; RMSEA = Root Mean Square Error of Approximation; IC90% = Intervalo de Confianza 90% RMSEA; P-close = Significación de la prueba de ajuste (RMSEA < 0.05); TLI = Tucker-Lewis Index; CFI = Comparative Fit Index.
Fuente: elaboración propia.

entre ambas submuestras son mínimas (53.98 % los varones frente a 46.02 % las mujeres en la modalidad de puntuación ordinal y 53.74 % los varones y 46.26 % las mujeres en la modalidad binaria). El valor RMSEA es razonable (0.072 en la modalidad ordinal y 0.053 en la binaria), y el resto de índices denotan un ajuste satisfactorio (TLI = 0.97; CFI = 0.98; TLI = 0.99; CFI = 0.99 en las modalidades ordinal y binaria, respectivamente). Estos resultados prestan apoyo a la equivalencia de ambas matrices de correlación.

Similares resultados se han obtenido al comparar las correlaciones de las puntuaciones obtenidas en los tres factores de la escala por ambas submuestras. El valor de χ^2 global ha resultado no significativo ($\chi^2_{(6)} = 2.02, p = 0.92$ en la modalidad ordinal; $\chi^2_{(6)} = 6.08, p = 0.41$ en la modalidad binaria). La contribución a χ^2 ha sido similar en ambas submuestras (52.4 % en los varones y 47.6 % en las mujeres en la modalidad ordinal; 54.33 % en los varones y 45.67 % en las mujeres en la modalidad binaria). Todos los ín-

dices de ajuste parcial han resultado altamente satisfactorios.

El segundo procedimiento (Cohen, Cohen, West & Aiken, 2003) consistió en la comparación individual de los pares de correlaciones (153 en el caso de los ítems y 3 en el de los factores). En la modalidad ordinal, únicamente 3 de las 153 comparaciones (1.96 %) han obtenido un valor z de Fisher igual o superior al valor crítico de 2.58, por lo que estos resultados refrendan los obtenidos mediante la metodología de ecuaciones estructurales (véanse Apéndices A y B). Más cuestionables han sido los resultados del análisis de las matrices en la modalidad binaria: de las 153 comparaciones, 19 (12.4 %) han resultado en un valor $z > 2.58$. Es de señalar que 10 de ellas corresponden al ítem 17 ('Tiene dificultades para guardar turno').

Ninguna de las correlaciones entre las puntuaciones obtenidas en los tres factores ha resultado en diferencias significativas con $p < 0.01$. En consecuencia, podemos concluir que niños y niñas tienen estructuras de correlación estadísticamente equiparables, tanto considerando los 18 ítems individuales como los tres factores de la escala.

Análisis RLO de los síntomas de déficit de atención

En la Tabla 4 pueden observarse los resultados del análisis RLO de los síntomas correspondientes a la subescala de Desatención. El primero de los pasos corresponde a la puntuación total del modelo (i. e., variable condicional); el segundo, al DIF uniforme (i. e., variable de agrupación); y el tercero al DIF uniforme y no uniforme (i. e., término de interacción). Para cada paso se proporcionan los valores de R^2 , junto con el tamaño del efecto entre cada par de pasos. El rango de los valores $\lambda\chi^2$ es de 4.18 (de 0.01 a 4.19) para la modalidad de calificación ordinal y de 2.89 (de 0.19 a 3.07) para la modalidad de calificación binaria. Ninguno de ellos, en consecuencia, supera el valor crítico de 9.21 necesario para considerar un incremento en χ^2 significativo al nivel de confianza del 99 %. El tamaño del efecto abarca un rango de -0.004 a 0.01 en la calificación ordinal y de -0.004 a 0.022 en la calificación bina-

ria. Se concluye, por tanto, que el tamaño del efecto es insignificante.

Análisis RLO de los síntomas de hiperactividad

En la Tabla 5 pueden observarse los resultados del análisis RLO de los síntomas correspondientes a la subescala de Hiperactividad. El rango es de 10.39 (de 0.67 a 11.06) en la modalidad ordinal y de 14.221 (de 0.769 a 14.99) en la binaria. En la ordinal, se observa un único incremento significativo del χ^2 en el paso 3 con respecto al χ^2 en el paso 1 ($\lambda\chi^2_{(2)} = 11.06$; $p = 0.004$) en el ítem 13 ('Tiene dificultades para jugar o dedicarse tranquilamente a actividades de ocio'). En la binaria, el único valor significativo de la diferencia entre los pasos 3 y 1 corresponde al ítem 15 ('Habla en exceso'), cuyo valor $\lambda\chi^2_{(2)}$ es de 14.99 ($p = 0.001$). No se puede, sin embargo, admitir la presencia de DIF en dichos ítems, toda vez que el tamaño del efecto de Zumbo-Thomas (χR^2) no alcanza el valor crítico de 0.130 (Zumbo, 1999). El tamaño del efecto abarca un rango de -0.001 a 0.021 en la calificación ordinal y de -0.001 a 0.042 en la calificación binaria y es, como en el caso anterior, insignificante. Es preciso señalar que en la modalidad de calificación binaria no ha sido posible realizar el análisis RLO del ítem 13 ('Tiene dificultades para jugar o dedicarse tranquilamente a actividades de ocio'): dicho ítem alcanza únicamente el 5.70 % de respuestas positivas en la modalidad de calificación binaria, con lo que la probabilidad de contestar positivamente es muy baja y las puntuaciones en este ítem no proporcionan suficiente variabilidad para llevar a cabo el análisis.

Análisis RLO de los síntomas de impulsividad

En la Tabla 6 se presentan los resultados del análisis RLO de los tres síntomas correspondientes a la subescala de Impulsividad. El rango es de 0.67 (de 0.06 a 0.73) en la modalidad ordinal y de 2.81 (de 0.77 a 3.58) en la binaria. El tamaño del efecto abarca un rango de 0.0000 a 0.0001 en la califica-

FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM EN LA EVALUACIÓN DE TDAH
EN FUNCIÓN DEL GÉNERO Y EL FORMATO DE CALIFICACIÓN

TABLA 4
Análisis DIF de los síntomas de desatención

	DESATENCIÓN: CALIF. ORDINAL	Pasos del análisis			$\lambda\chi^2(2)$		R ²			χ^2R^2			Fiab.
		$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$	P3-P1	p	Paso 1	Paso 2	Paso 3	P3-P1	P2-P1	P3-P2	
i1	no atención trabajo	483.51	487.09	487.71	4.19	0.123	0.63	0.635	0.639	0.01	0.005	0.005	0.752
i2	no atención juego	298.56	298.89	301.84	3.27	0.195	0.488	0.488	0.497	0.009	0	0.009	0.74
i3	no escucha	374.5	374.5	374.79	0.29	0.863	0.527	0.527	0.53	0.003	0	0.003	0.648
i4	no sigue instrucciones	4440.27	455.6	445.95	1.68	0.433	0.599	0.602	0.604	0.005	0.003	0.002	0.721
i5	dificultades organizar tareas	480.24	480.47	481.48	1.24	0.538	0.634	0.634	0.631	-0.003	0	-0.003	0.787
i6	no esfuerzo mental sostenido	432.3	432.3	432.31	0.01	0.995	0.597	0.597	0.596	0	0	0	0.778
i7	extravía objetos	239.56	240.15	240.97	1.41	0.493	0.395	0.397	0.4	0.005	0.002	0.003	0.563
i8	distracción fácil	553.16	553.5	555.04	1.88	0.391	0.702	0.702	0.698	-0.004	0	-0.004	0.774
i9	descuidado	461.8	461.94	464.23	2.44	0.296	0.617	0.617	0.614	-0.002	0.001	-0.003	0.775
	DESATENCIÓN: CALIF. BINARIA	Pasos del análisis			$\lambda\chi^2(2)$		R ²			χ^2R^2			
		$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$	P3-P1	p	Paso 1	Paso 2	Paso 3	P3-P1	P2-P1	P3-P2	
i1	no atención trabajo	263.21	263.33	264.76	1.56	0.459	0.548	0.549	0.549	0.001	0.001	0	---
i2	no atención juego	184.12	184.23	184.54	0.42	0.812	0.456	0.457	0.456	0	0.001	-0.001	---
i3	no escucha	201.49	201.59	204.56	3.07	0.215	0.441	0.442	0.461	0.02	0.001	0.019	---
i4	no sigue instrucciones	221.42	222.22	222.36	0.93	0.628	0.504	0.509	0.512	0.009	0.005	0.003	---
i5	dificultades organizar tareas	269.07	269.66	270.99	1.92	0.383	0.577	0.578	0.575	-0.002	0.001	-0.003	---
i6	no esfuerzo mental sostenido	203.51	203.51	206.33	2.82	0.244	0.46	0.46	0.48	0.02	0	0.02	---
i7	extravía objetos	146.03	146.47	147.26	1.23	0.54	0.374	0.378	0.397	0.022	0.004	0.019	---
i8	distracción fácil	325.32	325.52	325.99	0.67	0.717	0.65	0.65	0.646	-0.004	0	-0.004	---
i9	descuidado	242.65	242.65	242.83	0.19	0.912	0.547	0.547	0.547	0	0	0	---

Nota. Fiab. = Estimaciones de la fiabilidad de los ítems.
Fuente: elaboración propia.

TABLA 5
Análisis DIF de los síntomas de hiperactividad

	HIPERACTIVIDAD: CALIF. ORDINAL	Pasos del análisis			$\lambda\chi^2(2)$		R ²			χ^2R^2			Fiab.
		$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$	P3-P1	p	Paso 1	Paso 2	Paso 3	P3-P1	P2-P1	P3-P2	
i10	mueve manos y pies	721.77	721.77	722.74	0.98	0.614	0.76	0.76	0.759	-0.001	0	-0.001	0.799
i11	abandona asiento	632.16	632.19	634.5	2.34	0.31	0.704	0.704	0.703	-0.001	0	-0.001	0.786
i12	corre o salta	618.66	618.82	621.83	3.18	0.204	0.714	0.714	0.716	0.002	0	0.002	0.865
i13	dificultades jugar	154.65	162.34	165.71	11.06	0.004	0.346	0.366	0.357	0.012	0.021	-0.009	0.817
i14	está en marcha	613.23	614.23	614.61	1.38	0.502	0.714	0.714	0.716	0.001	0	0.002	0.82
i15	habla en exceso	377.34	377.8	378.01	0.67	0.715	0.491	0.492	0.494	0.003	0.001	0.002	0.587
	HIPERACTIVIDAD: CALIF. BINARIA	Pasos del análisis			$\lambda\chi^2(2)$		R ²			χ^2R^2			
		$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$	P3-P1	p	Paso 1	Paso 2	Paso 3	P3-P1	P2-P1	P3-P2	
i10	mueve manos y pies	415.75	421.78	422.09	6.34	0.042	0.691	0.701	0.703	0.012	0.009	0.003	---
i11	abandona asiento	393.71	395.04	395.73	2.02	0.364	0.663	0.665	0.667	0.004	0.003	0.001	---
i12	corre o salta	334.42	334.78	336.24	1.82	0.402	0.644	0.645	0.652	0.009	0.001	0.008	---
i13	dificultades jugar	---	---	---	---	---	---	---	---	---	---	---	---
i14	está en marcha	345.29	350.26	350.37	5.07	0.079	0.635	0.646	0.645	0.01	0.011	-0.001	---
i15	habla en exceso	284.92	291.49	299.91	14.99	0.001	0.507	0.52	0.55	0.042	0.013	0.029	---

Nota. Fiab. = Estimaciones de la fiabilidad de los ítems; en **negrita**, valores de $\lambda\chi^2(2)$ significativos.
Fuente: elaboración propia.

TABLA 6
Análisis DIF de los síntomas de impulsividad

IMPULSIVIDAD: CALIF. ORDINAL	Pasos del análisis			$\lambda\chi^2(2)$ P3-P1	p	R2			χ^2R^2			Fiab.
	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$			Paso 1	Paso 2	Paso 3	P3-P1	P2-P1	P3-P2	
i16 respuestas rápidas	453.13	453.33	453.35	0.22	0.897	0.602	0.603	0.603	0	0	0	0.521
i17 no guarda turno	648.65	648.7	648.71	0.06	0.969	0.739	0.739	0.739	0	0	0	0.753
i18 no esfuerzo mental sost.	684	684.72	684.73	0.73	0.695	0.757	0.758	0.758	0	0	0	0.792
IMPULSIVIDAD: CALIF. BINARIA	Pasos del análisis			$\lambda\chi^2(2)$ P3-P1	p	R2			χ^2R^2			Fiab.
	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(3)$			Paso 1	Paso 2	Paso 3	P3-P1	P2-P1	P3-P2	
i16 respuestas rápidas	160.22	160.99	160.99	0.77	0.681	0.404	0.406	0.406	0.002	0.001	0	---
i17 no guarda turno	255.04	255.14	258.62	3.58	0.167	0.517	0.517	0.54	0.023	0	0.024	---
i18 interrumpe	234.6	234.63	236.03	1.43	0.49	0.472	0.472	0.48	0.007	0	0.007	---

Nota. Fiab. = Estimaciones de la fiabilidad de los ítems.
Fuente: elaboración propia.

ción ordinal y de 0.000 a 0.024 en la calificación binaria. En consecuencia, ninguno de los tres ítems presenta DIF.

Podemos determinar, a la luz de los resultados anteriores, que únicamente los ítems 13 y 15 estarían en riesgo de presentar DIF.

Discusión

En este estudio se intentó contrastar el ajuste a los datos empíricos de distintos modelos relativos a la estructura de la sintomatología del trastorno por déficit de atención con hiperactividad (TDAH) propuesta por el sistema clasificatorio DSM-IV-TR (APA, 2002). De los cinco modelos sometidos a prueba, el que ha presentado un mejor ajuste es el de tres factores correlacionados (Desatención, Hiperactividad e Impulsividad). También presentó un ajuste aceptable el de dos factores correlacionados (Desatención e Hiperactividad/Impulsividad). Si bien la mayor parte de los estudios se han decantado por este segundo modelo, en esta investigación se optó por el primero, debido a su mejor ajuste a los datos.

Por otra parte, se analizó la presencia de DIF en los 18 ítems que conforman la escala, atendiendo al género de los sujetos evaluados, con el fin de obtener evidencia empírica de la invarianza de los síntomas en niños y niñas, tal como se deduce de los criterios diagnósticos del DSM-IV-TR. Se comprobó, finalmente, si dicha invarianza se mantiene

independientemente del tipo de formato de calificación (ordinal o binario) utilizado en el diagnóstico. Utilizar ambos formatos obedeció a que el primero suele utilizarse en contextos de investigación y educativos y el segundo, en situaciones de diagnóstico. De los hallazgos más significativos del presente estudio se pueden extraer las conclusiones que se apuntan a continuación.

En la subescala Déficit de Atención, en la modalidad de calificación ordinal, ninguno de los ítems supera el valor crítico necesario para considerar un incremento en χ^2 significativo al nivel de confianza del 99 %. El tamaño del efecto abarca un rango muy reducido en la calificación ordinal y lo mismo ocurre en la calificación binaria (i. e., presencia/ausencia del síntoma evaluado). Se concluyó, por tanto, que el tamaño del efecto es insignificante y, en consecuencia, ninguno de los ítems presenta DIF en ninguna de las modalidades de calificación.

En la subescala Hiperactividad, en el formato de calificación ordinal únicamente el ítem 13 (“Tiene dificultades para jugar o dedicarse tranquilamente a actividades de ocio”) es sospechoso de presentar DIF. En la calificación binaria, lo es el ítem 15 (“Habla en exceso”). No podemos sin embargo admitir la presencia de DIF en dichos ítems, toda vez que el tamaño del efecto de Zumbo-Thomas (χ^2R^2) no alcanza el valor crítico de 0.13.

Se debe poner de manifiesto que en la modalidad de calificación binaria no fue posible realizar el análisis mediante regresión logística ordinal del

ítem 13 (“Tiene dificultades para jugar o dedicarse tranquilamente a actividades de ocio”): dicho ítem alcanza únicamente el 5.7 % de respuestas positivas en la modalidad de calificación binaria, con lo que la probabilidad de contestar positivamente es muy baja y las puntuaciones en este ítem no proporcionan suficiente variabilidad para llevar a cabo el análisis. Este hecho aconseja que, en futuras investigaciones sobre este problema, se trate de incluir en la muestra sujetos (sería aconsejable que provinieran de poblaciones clínicas) que tengan una mayor probabilidad de ser calificados positivamente en dicho ítem.

En la subescala Impulsividad, el tamaño del efecto (χR^2) abarca un rango muy reducido, tanto en la calificación ordinal como en la calificación binaria. En consecuencia, ninguno de los tres ítems de esta subescala presenta DIF.

Se puede determinar, a la luz de los resultados anteriores, que únicamente dos de los 18 ítems estarían en riesgo de presentar DIF. En consecuencia, los resultados del presente estudio proporcionan apoyo empírico a la no existencia de DIF en función del género de los niños evaluados, ni en la modalidad de calificación ordinal (escala de 1 a 4) ni en la modalidad de calificación binaria (0-1) de los ítems. Estos resultados refrendan el hecho de que en el *DSM-IV-TR* no se establezcan criterios diferenciales para el diagnóstico del TDAH en niños y niñas, y concuerdan, en términos generales, con los resultados de investigaciones realizadas en otros ámbitos geográficos (p. ej., Burns, Walsh, Gomez & Hafetz, 2006; Gomez, 2007; Reid et al., 1998) respecto a la invarianza de los síntomas TDAH atendiendo al género de los niños evaluados.

Entre las limitaciones del presente estudio cabe señalar las siguientes: en primer lugar, el carácter no probabilístico de la muestra utilizada imposibilita la generalización de los resultados. Convendría, en este sentido, utilizar muestras probabilísticas en ulteriores investigaciones sobre el problema estudiado. En segundo lugar, sería deseable utilizar niños y niñas provenientes de población clínica, con el objeto de averiguar si la invarianza por género se mantiene. En tercer lugar, el uso de muestras clínicas posibilitaría que todos los ítems pudieran

someterse al análisis con regresión logística (como se recordará, no ha sido posible realizar dicho análisis sobre el ítem “Tiene dificultades para jugar o dedicarse tranquilamente a actividades de ocio” en el formato de calificación binaria, debido a la carencia de suficiente variabilidad en las calificaciones).

Referencias

- Abell, N., Springer, D. W. & Kamata, A. (2009). *Developing and validating rapid assessment instruments*. New York: Oxford University Press.
- Amador, J. A., Forns, M., Guàrdia, J. & Peró, M. (2005). Utilidad diagnóstica del cuestionario TDAH y del perfil de atención para discriminar entre niños con trastorno por déficit de atención con hiperactividad, trastornos del aprendizaje y controles. *Anuario de Psicología*, 36(2), 211-224.
- Amador, J. A., Forns, M., Guàrdia, J. & Peró, M. (2006a). DSM-IV Attention Deficit Hyperactivity Disorder Symptoms: Agreement between informants in prevalence and factor structure at different ages. *Journal of Psychopathology and Behavioral Assessment*, 28(1), 23-32.
- Amador, J. A., Forns, M., Guàrdia, J. & Peró, M. (2006b). Estructura factorial y datos descriptivos del perfil de atención y del cuestionario TDAH para niños en edad escolar. *Psicothema*, 18(4), 696-703.
- Amador, J. A., Forns, M. & Martorell, B. (2001). Trastorno por déficit de atención con hiperactividad: análisis evolutivo y consistencia entre informantes. *Anuario de Psicología*, 3(1), 51-66.
- American Academy of Child and Adolescent Psychiatry. (1997). Practice parameters for the assessment and treatment of children, adolescent and adults with attention deficit hyperactivity disorder (AACAP official action). *Journal of American Academy of Child and Adolescent Psychiatry*, 36(10), 85S-121S.
- American Psychiatric Association. (2002). *DSM-IV-TR. Manual diagnóstico y estadístico de los trastornos mentales* (ed. rev.). Barcelona: Masson.
- Arnold, L. D. (1996). Sex differences in ADHD: Conference summary. *Journal of Abnormal Child Psychology*, 24(5), 555-570.
- Balluerka, M. N., Gómez, J., Stock, W. A. & Caterino, L. C. (2000). Características psicométricas de las

- versiones americana y española de la escala TDAH (Trastorno por Déficit de Atención con Hiperactividad): un estudio comparativo. *Psicothema*, 12(4), 629-634.
- Barkley, R. A. (2006). *Attention-Deficit Hyperactivity Disorder: A handbook for diagnosis and treatment* (3a ed.). New York: Guilford.
- Barkley, R. A. & Murphy, K. (2006). *Attention-Deficit Hyperactivity Disorder. A clinical workbook*. New York: Guilford.
- Bauermeister, J. J., Berrios, V., Jimenez, A. L., Acevedo, L. & Gordon, M. (1990). Some issues and instruments for the assessment of attention-deficit hyperactivity disorder in Puerto Rican children. *Journal of Clinical Child Psychology*, 19(1), 9-16.
- Benjumea, P. & Mojarro, M. D. (1993). Trastornos hiperkinéticos: un estudio epidemiológico en doble fase de una población sevillana. *Anales de Psiquiatría*, 9(7), 306-311.
- Burns, G. L., Walsh, J. A. & Gomez, R. (2003). Convergent and discriminant validity of trait and source effects in ADHD-inattention and hyperactivity-impulsivity measures across a 3-month interval. *Journal of Abnormal Child Psychology*, 31(5), 529-541.
- Burns, G. L., Walsh, J. A., Gomez, R. & Hafetz, N. (2006). Measurement and structural invariance of parent ratings of ADHD and ODD symptoms across gender for American and Malaysian children. *Psychological Assessment*, 18(4), 452-457.
- Cardo, E. & Servera, M. (2005). Prevalencia del trastorno de déficit de atención e hiperactividad. *Revista de Neurología*, 40(Supl. 1), S11-S15.
- Cardo, E., Servera, M. & Llobera, J. (2007). Estimación de la prevalencia del trastorno por déficit de atención e hiperactividad en población normal de la isla de Mallorca. *Revista de Neurología*, 44(1), 10-14.
- Clauser, B. E. & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31-44.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Cronbach, L. J., Schönemann, P. & McKie, D. (1965). Alpha coefficients for stratified- parallel tests. *Educational and Psychological Measurement*, 25(2), 291-312.
- DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M. J. & McGoey, K. E. (1998). Parent rating of attention deficit hyperactivity disorder symptoms: Factor structure and normative data. *Journal of Psychopathology and Behavioral Assessment*, 20(1), 83-102.
- DuPaul, G. J., Power, T. J., Anastopoulos, A. D., Reid, R., McGoey, K. E. & Ikeda, M. J. (1997). Teacher rating of attention deficit hyperactivity disorder symptoms: Factor structure and normative data. *Psychological Assessment*, 9(4), 436-443.
- Elosua, P. & Zumbo, B. D. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, 20(4), 896-901.
- Ferreres, D., Fidalgo, A. M. & Muñiz, J. (2000). Detección del funcionamiento diferencial de los ítems no uniforme: comparación de los métodos Mantel-Haenszel y regresión logística. *Psicothema*, 12(Supl. 2), 220-225.
- Gadow, K. D. & Sprafkin, J. (1997). *Symptom Checklist-4 Manual*. Stony Brook, NY: Checkmate Plus.
- Gelin, M. N., Carleton, B. C., Smith, A. & Zumbo, B. D. (2004). The dimensionality and gender differential item functioning of the Mini Asthma Quality of Life Questionnaire. *Social Indicators Research*, 68(1), 91-105.
- Gelin, N. M. & Zumbo, B. D. (2003). Differential item functioning results may change depending on how an item is scored: An illustration with the Center for Epidemiologic Studies Depression Scale. *Educational and Psychological Measurement*, 63(1), 65-74.
- Gomez, R. (2007). Testing gender differential item functioning for ordinal and binary scored parent rated ADHD symptoms. *Personality and Individual Differences*, 42(4), 733-742.
- Gomez, R., Burns, G., Walsh, J. & Hafetz, N. (2005). A multitrait-multisource confirmatory factor analytic approach to the construct validity of ADHD and ODD Rating Scales with Malaysian children. *Journal of Abnormal Child Psychology*, 33(2), 241-254.
- Gomez, R., Harvey, J., Quick, C., Scharer, I. & Harris, G. (1999). DSM-IV AD/HD: Confirmatory factor models, prevalence, and gender and age differences

- based on parent and teacher ratings of Australian primary school children. *Journal of Child Psychology and Psychiatry*, 40(2), 265-274.
- Hidalgo, M. D. & Gómez, J. (2000). Comparación de la eficacia de la regresión logística polinómica y análisis discriminante logístico en la detección del DIF no uniforme. *Psicothema*, 12(Supl. 2), 298-300.
- Hidalgo, M. D. & Gómez, J. (2003). Test purification and the evaluation of differential item functioning with multinomial logistic regression. *European Journal of Psychological Assessment*, 19(1), 1-11.
- Hidalgo, M. D., Gómez, J. & Padilla, J. L. (2005). Regresión logística: alternativas de análisis en la detección del funcionamiento diferencial del ítem. *Psicothema*, 17(3), 509-515.
- Hidalgo, M. D. & López, J. A. (2004). DIF detection and effect size: A comparison between logistic regression and Mantel-Haenszel variation. *Educational and Psychological Measurement*, 64(6), 903-915.
- Hinshaw, S. P., Han, S. S., Erhardt, D. & Huber, A. (1992). Internalizing and externalizing behavior problems in preschool children: Correspondence among parent and teachers rating and behavior observation. *Journal of Clinical Psychology*, 21(2), 143-150.
- Hosmer, D. W. & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: Wiley.
- Jodoin, M. G. & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14(4), 329-349.
- Jöreskog, K. & Sörbom, D. (2008). *LISREL 8.80 User's Guide*. Lincolnwood, IL: Scientific Software International.
- Kline, R. B. (2010). *Principles and practice of structural equation modeling*. New York: Guilford.
- LimeSurvey. (2009). LimeSurvey, v. 1.87 [Software de cómputo]. Disponible en <http://www.limesurvey.org>
- Muthén, B. & Muthén, L. (2011). MPlus, v. 6.1 [Software de cómputo]. Los Angeles, CA: StatModel.
- Polanczyk, G., De Dima, M. S., Horta, B. L., Biederman, J. & Rohde, L. A. (2007). The worldwide prevalence of ADHD: A systematic review and meta-regression analysis. *American Journal of Psychiatry*, 164(6), 942-948.
- Reid, R., DuPaul, G. J., Power, T. J., Anastopoulos, A., Rogers-Adkinson, D., Noll, M. B., et al. (1998). Assessing culturally different students for attention deficit hyperactivity disorder using behavior rating scales. *Journal of Abnormal Child Psychology*, 26(3), 187-198.
- Rodríguez, L., López-Villalobos, J. A., Garrido, M., Sacristán, A. M., Martínez, M. T. & Ruiz, F. (2009). Estudio psicométrico-clínico de prevalencia y comorbilidad del trastorno por déficit de atención con hiperactividad en Castilla y León (España). *Revista de Pediatría y Atención Primaria*, 11(42), 251-270.
- Scientific Software International. (2008). LISREL (v. 8.8) [Programa Informático]. Lincolnwood, IL: Scientific Software International.
- Servera, M. & Cardo, E. (2007). ADHD Rating Scale-IV en una muestra escolar española: datos normativos y consistencia interna para maestros, padres y madres. *Revista de Neurología*, 45(7), 393-399.
- Spencer, T. J., Biederman, J. & Mick, E. (2007). Attention-deficit/hyperactivity disorder: Diagnosis, lifespan, comorbidities, and neurobiology. *Journal of Pediatric Psychology*, 32(6), 631-642.
- Swaminathan, H. & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370.
- Swanson, J. (2010). *The SNAP-IV Rating Scale*. Recuperado el 25 de febrero de 2010, de <http://www.adhd.net>
- Von Eye, A. & Mun, E. Y. (2005). *Analyzing rater agreement*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wolraich, M., Lambert, E. W., Doffing, M. A., Bickman, L., Simmons, T. & Worley, K. A. (2003). Psychometric properties of the Vanderbilt ADHD diagnostic parent rating scale in a referred population. *Journal of Pediatric Psychology*, 28(8), 559-567.
- Whitmore, M. L. & Schumacker, R. E. (1999). A comparison of logistic regression and analysis of variance differential item functioning detection methods. *Educational and Psychological Measurement*, 59(6), 910-927.

Zumbo, B. D. (1999). *A handbook on the theory and methods of Differential Item Functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) items scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zumbo, B. D., Gadermann, A. M. & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21-29.

APÉNDICE A
Valores z y niveles de significación (modalidad ordinal)

TABLA A1
Valores z y niveles de significación (modalidad ordinal)

	i01	i02	i03	i04	i05	i06	i07	i08	i09	i10	i11	i12	i13	i14	i15	i16	i17	i18	DA	HI	IM
i01		0.011	0.361	0.032	0.336	0.856	0.004	0.614	0.556	0.221	0.112	0.964	0.141	0.541	0.554	0.493	0.989	0.167			
i02	2.54		0.034	0.001	0.046	0.211	0.000	0.969	0.300	0.342	0.933	0.003	0.131	0.326	0.903	0.442	0.001	0.437			
i03	0.91	2.12		0.191	0.603	0.042	0.066	0.284	0.333	0.353	0.911	0.016	0.282	0.254	0.474	0.336	0.819	0.179			
i04	2.14	3.20	1.31		0.023	0.002	0.005	0.002	0.083	0.035	0.160	0.074	0.013	0.005	0.043	0.311	0.001	0.324			
i05	-0.96	2.00	0.52	2.28		0.568	0.256	0.194	0.144	0.884	0.669	0.857	0.558	0.558	0.989	0.843	0.412	0.572			
i06	0.18	1.25	2.03	3.06	0.57		0.269	0.166	0.035	0.941	0.337	0.446	0.503	0.077	0.670	0.830	0.096	0.988			
i07	2.87	5.09	1.84	2.82	-1.14	1.10		0.004	0.000	0.091	0.263	0.005	0.000	0.002	0.043	0.805	0.063	0.956			
i08	0.50	-0.04	1.07	3.10	1.30	1.38	2.89		0.036	0.185	0.171	0.734	0.586	0.757	0.424	0.285	0.626	0.101			
i09	-0.59	1.04	0.97	1.73	1.46	2.11	5.58	2.10		0.503	0.686	0.963	0.172	0.570	0.490	0.057	0.875	0.103			
i10	-1.22	0.95	0.93	2.11	-0.15	0.07	1.69	-1.33	0.67		0.056	0.962	0.089	0.429	0.844	0.769	0.181	0.198			
i11	-1.59	-0.08	-0.11	1.40	0.43	0.96	1.12	-1.37	0.40	-1.91		0.584	0.125	0.082	0.797	0.344	0.542	0.819			
i12	0.04	3.00	2.41	1.78	-0.18	0.76	2.78	-0.34	0.05	0.05	0.55		0.784	1.000	0.841	0.603	0.234	0.593			
i13	1.47	1.51	-1.07	2.49	0.59	0.67	4.50	-0.54	-1.37	-1.70	-1.53	0.27		0.046	0.978	0.116	0.183	0.386			
i14	-0.61	0.98	1.14	2.82	0.59	1.77	3.12	0.31	0.57	-0.79	-1.74	0.00	-2.00		0.444	0.381	0.486	0.016			
i15	-0.59	-0.12	-0.72	2.02	-0.01	-0.43	2.03	-0.80	-0.69	-0.20	-0.26	0.20	0.03	-0.76		0.012	0.132	0.672			
i16	0.69	-0.77	-0.96	1.01	-0.20	0.22	-0.25	-1.07	-1.91	-0.29	-0.95	-0.52	1.57	-0.88	-2.52		0.090	0.178			
i17	0.01	3.36	0.23	3.36	0.82	1.67	1.86	0.49	0.16	1.34	0.61	1.19	1.33	-0.70	-1.51	-1.69		0.128			
i18	-1.38	0.78	-1.34	0.99	-0.56	-0.02	0.06	-1.64	-1.63	-1.29	0.23	-0.54	0.87	-2.42	-0.42	-1.35	-1.52				
DA																			0.260	0.889	
HI																			1.127	0.691	
IM																			0.139	0.397	

Nota. Valores z (mitad inferior) y niveles de significación (mitad superior); DA = Déficit de Atención; HI = Hiperactividad; IM = Impulsividad.
Fuente: elaboración propia.

APÉNDICE B

Valores z y niveles de significación (modalidad binaria)

TABLA B1
Valores z y niveles de significación (modalidad binaria)

	i01	i02	i03	i04	i05	i06	i07	i08	i09	i10	i11	i12	i13	i14	i15	i16	i17	i18	DA	HI	IM
i01	0.031	0.242	0.013	0.015	0.013	0.724	0.000	0.348	0.267	0.001	0.009	0.001	0.002	0.123	0.428	0.026	0.080	0.038			
i02	2.16		0.041	0.000	0.012	0.000	0.000	0.000	0.000	0.277	0.420	0.011	0.000	0.975	0.094	0.469	0.000	0.683			
i03	1.17	2.05		0.442	0.088	0.005	0.002	0.844	0.216	0.316	0.060	0.046	0.026	0.026	0.009	0.466	0.000	0.023			
i04	2.42	5.18	-0.77		0.264	0.012	0.000	0.049	0.000	0.190	0.003	0.302	0.000	0.123	0.082	0.470	0.020	0.828			
i05	2.48	2.52	1.71	1.12		0.122	0.000	0.000	0.000	0.490	0.114	0.143	0.000	0.535	0.012	0.005	0.527	0.003			
i06	-0.35	5.13	2.83	2.50	1.55		0.000	0.514	0.003	0.298	0.811	0.755	0.000	0.347	0.110	0.334	0.000	0.932			
i07	6.04	6.29	3.03	4.25	-4.67	-6.10		0.003	0.005	0.527	0.061	0.032	0.000	0.247	0.072	0.178	0.000	0.029			
i08	-0.94	4.78	-0.20	1.97	4.08	0.65	3.00		0.094	0.032	0.526	0.097	0.265	0.185	0.814	0.005	0.000	0.007			
i09	1.11	5.32	-1.24	3.85	4.62	2.94	2.79	1.68		0.059	0.000	0.000	0.431	0.003	1.000	0.986	0.006	0.001			
i10	-3.39	-1.09	-1.00	-1.31	0.69	-1.04	0.63	-2.14	1.89		0.006	0.331	0.000	0.522	0.144	0.824	0.000	0.012			
i11	-2.61	-0.81	-1.88	3.02	1.58	0.24	1.88	-0.63	4.79	-2.77		0.003	0.333	0.024	0.000	0.886	0.005	0.284			
i12	-3.39	2.55	2.00	1.03	1.46	-0.31	2.15	1.66	7.70	0.97	2.97		0.711	0.132	0.440	0.064	0.001	0.774			
i13	3.14	4.39	2.22	6.01	3.56	4.29	6.03	1.11	0.79	4.00	-0.97	-0.37		0.002	0.175	0.000	0.000	0.322			
i14	-1.54	-0.03	-2.22	1.54	0.62	0.94	1.16	1.33	2.95	-0.64	-2.25	1.51	3.14		0.041	0.312	0.051	0.624			
i15	-0.79	-1.67	-2.61	1.74	2.52	1.60	1.80	0.23	0.00	-1.46	-3.75	-0.77	-1.36	-2.05		0.199	0.513	0.000			
i16	-2.22	0.72	-0.73	-0.72	-2.81	0.97	1.35	-2.81	-0.02	0.22	0.14	1.85	4.24	1.01	-1.28		0.646	0.336			
i17	10.75	6.24	3.94	2.33	0.63	4.33	3.49	3.92	2.73	4.66	2.83	3.31	6.74	1.95	-0.65	0.46		0.194			
i18	-2.08	-0.41	-2.28	-0.22	-2.99	-0.09	2.19	-2.71	-3.26	-2.52	1.07	-0.29	0.99	-0.49	-3.89	-0.96	-1.30				
DA																			0.331	0.857	
HI																			0.97		0.078
IM																			-0.18	1.76	

Nota. Valores z (mitad inferior) y niveles de significación (mitad superior); DA = Déficit de Atención; HI = Hiperactividad; IM = Impulsividad. Fuente: elaboración propia.