

EXPLORACIÓN DE LA ESTRUCTURA FENOTÍPICA DEL TRASTORNO POR DÉFICIT DE ATENCIÓN CON HIPERACTIVIDAD (TDAH): SUBTIPOS FRENTE A CONTINUO DE GRAVEDAD

Víctor B. Arias¹, Jonatan Frutos², Henar Rodríguez² y Benito Arias²

¹Universidad de Salamanca; ²Universidad de Valladolid (España)

Resumen

El DSM-5 requiere para el diagnóstico del trastorno por déficit de atención con hiperactividad (TDAH) la presencia de al menos seis síntomas, dando lugar a tres subtipos o presentaciones. Sin embargo la validez de los subtipos sigue siendo objeto de debate. Nuestros objetivos fueron (a) investigar si el TDAH es mejor caracterizado como una tipología, un continuo o una mixtura de ambos, e (b) investigar si las presentaciones del TDAH propuestas por el DSM-5 pueden ser empíricamente validadas. Para ello, comparamos modelos factoriales, modelos de perfiles latentes y modelos factoriales mixtos estimados sobre los datos provenientes de 871 niños (de 5 a 14 años). El mejor modelo fue el de dos factores (déficit de atención e hiperactividad/impulsividad) y cuatro clases. Las clases no representaron grupos con perfiles sintomáticos cualitativamente distintos, sino una mixtura de distribuciones aproximadamente normales subyacentes a un mismo continuo de gravedad. Estos resultados no apoyan la categorización de los niños en subtipos o presentaciones, sino su evaluación en un continuo de gravedad cuyo extremo superior podría entenderse como la manifestación clínica del TDAH.

PALABRAS CLAVE: *TDAH, subtipos, presentaciones, validez, modelo factorial mixto.*

Abstract

DSM-5 requires for the diagnosis of attention deficit hyperactivity disorder (ADHD) the presence of at least six symptoms, resulting in three subtypes or presentations. However, the validity of the subtypes remains debated. Our objectives were (a) to investigate whether ADHD is best characterized as a typology, a continuum, or a mixture of both, and (b) to investigate whether ADHD presentations proposed by DSM-5 can be empirically validated. For this purpose, we compared confirmatory factor models, latent profile models and mixture factor models on data from 871 children (5 to 14 years old). The best model was that of two factors (attention deficit and hyperactivity/impulsivity) and four classes. The classes did not represent groups with qualitatively different symptomatic profiles, but a mixture of approximately normal distributions underlying the same severity continuum. These results do not support the categorization of children in subtypes or presentations, but their assessment in a

continuum of severity whose upper end could be understood as the clinical manifestation of ADHD.

KEY WORDS: *ADHD, subtypes, presentations, validity, mixture factor model.*

Introducción

El trastorno por déficit de atención con hiperactividad (TDAH) es un trastorno del neurodesarrollo caracterizado por síntomas persistentes de déficit de atención (DA), hiperactividad (HI) e impulsividad (IM) que son desadaptativos e incoherentes con el nivel de desarrollo del niño (Asociación Americana de Psiquiatría, 2014). Se estima que la prevalencia del TDAH se encuentra entre el 3% y el 7% de la población infantil, con variaciones en función del tipo de estudio y el procedimiento de evaluación utilizado (Cardo, Servera y Llobera, 2007; Polanczyk, de Lima, Lessa, Biederman y Rohde, 2007). Además, el TDAH tiende a presentar tasas de comorbilidad muy elevadas, así como una gran heterogeneidad en la manifestación de sus síntomas, tanto entre diferentes individuos como a lo largo de la edad (Nigg, Tannock y Rohde, 2010).

En sus versiones cuarta y quinta, el criterio B del Manual diagnóstico y estadístico de los trastornos mentales (DSM) requiere para el diagnóstico de TDAH la presencia de, al menos, seis síntomas en cualquiera de los dos grupos (déficit de atención e hiperactividad/impulsividad), dando lugar a tres diagnósticos posibles: TDAH con presentación combinada, con predominio de déficit de atención o con predominio de hiperactividad/impulsividad. La adhesión a seis o más síntomas en al menos una de las listas conduce al diagnóstico de uno de los tres subtipos o presentaciones de TDAH (APA, 2014). Sin embargo, este procedimiento de categorización no ha estado exento de críticas. Por ejemplo, el algoritmo diagnóstico propuesto por el DSM puede conducir a situaciones donde un niño, cumpliendo diez síntomas de TDAH (cinco en cada grupo), no satisfaga el criterio B, mientras que un niño que cumpla seis en uno de los grupos sí sea candidato a diagnóstico (Hudziak, Achenbach, Althoff y Pine, 2007; Lubke *et al.*, 2007). Esto ha llevado al debate de si las diferencias fenotípicas observadas en niños con diagnóstico de TDAH son resultado de diferencias en la gravedad del trastorno a lo largo de un continuo subyacente, o si realmente reflejan subtipos cualitativamente distintos con aspectos etiológicos específicos (Gomez, Vance y Gomez, 2013; Lubke *et al.*, 2007; Pickles y Angold, 2003; Ranby *et al.*, 2012).

Los estudios dedicados a las manifestaciones conductuales del TDAH a menudo han aplicado análisis estadísticos que, explícita o implícitamente, han incorporado suposiciones sobre la estructura subyacente a los síntomas. Desde el DSM-III-R (APA, 1988), la estructura más frecuentemente atribuida al TDAH ha consistido en dos dimensiones correlacionadas (DA e HI/IM; Bauermeister, Canino, Polanczyk y Rohde, 2010; Willcutt *et al.*, 2012). Esta estructura, que denominaremos en adelante modelo oblicuo, ha ganado aceptación y presentado pruebas de estabilidad entre culturas y a lo largo de distintas etapas del desarrollo. Sin embargo, el modelo oblicuo presenta ciertas limitaciones psicométricas como, por ejemplo, correlaciones muy altas entre DA y HI/IM (Arias, Ponce y Núñez,

2016a; Willcutt *et al.*, 2012) que alteran la interpretación de las puntuaciones y ponen en duda la separabilidad empírica de las dos dimensiones como entidades cualitativamente diferentes. Además, es posible que varios de los 18 síntomas del TDAH presenten validez discriminante débil con respecto a su factor teórico (Arias, Ponce, Martínez-Molina, Arias y Núñez, 2016b). En resumen, aunque el modelo oblicuo haya demostrado estabilidad estructural, puede presentar dificultades en su especificación e interpretación, haciendo viables modelos alternativos. De estos, posiblemente sea el modelo bifactor el que más repercusión ha obtenido (Arias *et al.*, 2016a). El modelo bifactor propone la existencia de una dimensión general de TDAH, común a todos los síntomas y de, al menos, una dimensión residual de déficit de atención. En todo caso, tanto al modelo oblicuo como al bifactor subyace la hipótesis de que el TDAH se organiza en un continuo multidimensional, común a población general y clínica, donde las diferencias individuales se deben a diferencias en la gravedad del trastorno y no a la pertenencia a diferentes subtipos o presentaciones.

Sin embargo, el modelo dimensional por sí mismo aporta poca información útil para la clínica aplicada, donde es necesario establecer criterios y algoritmos diagnósticos claros que faciliten la toma de decisiones de clasificación correctas. En otras palabras, aunque la perspectiva dimensional del TDAH ha recibido amplio apoyo empírico (Willcutt *et al.*, 2012), presenta limitaciones a la hora de informar con claridad donde está el punto de corte apropiado para generar el diagnóstico. Una aproximación metodológica alternativa para investigar la estructura categorial del TDAH ha sido el análisis de clases latentes (ACL). El ACL (McCutcheon, 1987) es una aproximación útil para la validación de fenotipos psiquiátricos (Leoutsakos, Zandi, Bandeen-Roche y Lyketsos, 2010). Permite identificar perfiles sintomáticos sin la necesidad de establecer a priori umbrales diagnósticos (que en ocasiones pueden resultar en buena parte arbitrarios; Borsboom *et al.*, 2016). En el ACL se asume que las clases son heterogéneas entre sí y homogéneas internamente (es decir, las correlaciones observadas entre las puntuaciones de un sujeto en los síntomas individuales son interpretadas como resultado de su pertenencia a una clase concreta). El método implica la extracción del número de clases que presente mejor ajuste dada cierta distribución observada en las respuestas de los sujetos. Los estudios de TDAH que han utilizado análisis de clases latentes extraído entre dos y ocho clases (Bauermeister *et al.*, 2010). Como es lógico, tal diversidad de resultados dificulta llegar a conclusiones firmes acerca del número de clases o perfiles fenotípicos del TDAH. Una posible razón para explicar estos resultados es que el ACL no ofrece la posibilidad de modelar diferencias en gravedad intraclase, lo que podría dar lugar a la interpretación errónea de diferencias cuantitativas en gravedad como clases distintas (Lubke y Neale, 2008).

Dadas las limitaciones de los modelos dimensionales y de clases latentes, recientemente se ha propuesto el uso de un modelo híbrido conocido como modelo factorial mixto (MFM; Lubke y Muthén, 2005; Muthén y Shedden, 1999; Yung, 1997). El MFM permite analizar la estructura del trastorno combinando una perspectiva categorial con una dimensional en un solo modelo general. En consecuencia, el MFM permite que la estructura subyacente sea simultáneamente categorial (mediante la estimación de clases) y dimensional (mediante la inclusión

de un modelo factorial dentro de cada clase). En un primer paso, el MFM establece que la heterogeneidad observada en la manifestación de los síntomas obedece a la presencia de clases (o perfiles) distintos. Sin embargo, y a diferencia del ACL, se asume que las clases son internamente heterogéneas mediante la estimación de un modelo factorial confirmatorio o exploratorio para cada clase. En consecuencia, en un modelo MFM la variabilidad con que las personas presentan los síntomas se explica, por una parte, a través de su pertenencia a clases distintas (p. ej., afectados y no afectados), así como por la variabilidad intraclase (p. ej., se asume que los sujetos afectados pueden diferir entre sí en la gravedad con la que manifiestan los síntomas, como de hecho se observa en la realidad, dada la conocida heterogeneidad conductual del TDAH). Una ventaja adicional del MFM es que su ajuste puede ser comparado directamente con el de modelos factoriales y de clases latentes mediante algunos de los índices habitualmente utilizados en la literatura (Lubke y Neale, 2008).

Por sus características, el MFM parece un modelo útil para investigar la heterogeneidad del TDAH y contrastar empíricamente la existencia de subtipos o presentaciones empíricamente identificables. Sin embargo, hasta donde sabemos sólo tres estudios han empleado este modelo sobre los síntomas del TDAH. Lubke *et al.* (2007), sobre una muestra comunitaria de adolescentes, concluyeron que un modelo con dos factores (DA y HI/IM) y dos clases (probabilidad de adhesión alta y baja a los síntomas) representó de forma óptima la estructura subyacente a los síntomas. Gomez *et al.*, (2013) llegaron a idénticas conclusiones empleando una muestra clínica de niños y adolescentes, donde una clase de gravedad alta en todos los síntomas capturó la mayor parte de individuos diagnosticados con TDAH de tipo combinado. Por último, Ranby *et al* (2012), sobre una muestra de adultos que evaluaron en retrospectiva su propia conducta durante la niñez, también obtuvieron la combinación dos clases y dos factores como modelo más plausible. Cabe destacar que ninguno de los tres estudios, incluyendo el que empleó muestras clínicas (Gomez *et al.*, 2013), obtuvo resultados que apoyasen la presencia de subtipos o presentaciones, en cuanto que ninguna de las clases obtenidas representó un grupo de sujetos con elevaciones en un conjunto particular de síntomas (IN o HI/IM) y bajo en el otro.

La cuestión de si los constructos relativos a la psicopatología son mejor representados como categorías discretas o como dimensiones continuas es un problema de investigación importante en psicología clínica y psiquiatría (Borsboom *et al.*, 2016). En el caso del TDAH, los sistemas diagnósticos más utilizados como el DSM y la décima revisión de la Clasificación internacional de enfermedades (*International classification of diseases*, ICD-10; Organización Mundial de la Salud, 1991) se adhieren a una perspectiva categorial, donde el diagnóstico es de tipo discreto y se basa en la presencia de determinados patrones de síntomas. Uno de los riesgos de un sistema puramente categorial es que los puntos de corte para decidir si el trastorno está presente pueden ser en buena parte arbitrarios (Borsboom *et al.*, 2016). En el caso del TDAH, el criterio diagnóstico de seis o más síntomas presentes no ha recibido apoyo empírico suficiente que garantice su superioridad frente a otros puntos de corte (p. ej., cinco o siete síntomas). Por el contrario, ha recibido críticas relacionadas con la insensibilidad del criterio a las

diferencias de gravedad entre los propios síntomas y se ha propuesto la inclusión de un eje cuantitativo en la descripción diagnóstica que permita la presencia de diferencias de gravedad de acuerdo a variables relevantes como el sexo o la edad (Lubke *et al.*, 2009; Hudziak *et al.*, 1998). Por otra parte, la escasez de información sobre la validez de la estructura categorial del TDAH propuesta por el DSM puede deberse a la escasez de estudios al respecto: la mayoría de investigaciones sobre la estructura interna del TDAH se ha basado en modelos factoriales, donde el TDAH se representa como un rasgo puramente continuo cuyo límite superior constituye la manifestación clínica del trastorno (Willcutt *et al.*, 2012).

En el presente estudio nos propusimos investigar si el TDAH, evaluado por profesores en una muestra de niños proveniente de población general, es mejor caracterizado como una tipología, como un continuo, o como una mixtura de ambos. El primer objetivo específico fue replicar los resultados de estudios anteriores (Lubke *et al.*, 2007, Ranby *et al.*, 2012 y Gomez *et al.*, 2013) en una muestra española de niños con edades comprendidas entre los cinco y los 14 años. Dado el carácter eminentemente exploratorio de los modelos factoriales mixtos, la replicación mediante estudios independientes, especialmente en otras culturas e idiomas fuera del ámbito anglosajón, es fundamental para dotar de validez y generalizabilidad a los resultados. Por otra parte, los estudios descritos analizaron las puntuaciones de adolescentes, adultos o una combinación de niños y adolescentes; sin embargo, ninguno centró sus análisis en el tramo de edad que, según los criterios del DSM-IV-TR y DSM-5, es crítico para el diagnóstico del TDAH (entre nueve y 12 años). Por último, en este estudio hemos analizado evaluaciones de profesores. Los estudios arriba citados emplearon evaluaciones de padres (Gomez *et al.*, 2013, Lubke *et al.*, 2007) o autoinformes retrospectivos de adultos (Ranby *et al.*, 2012). No hay pues evidencia disponible de que los resultados obtenidos se puedan extrapolar a evaluaciones de profesores. Esta es una cuestión relevante, dado que cada tipo de informante proporciona información única sobre el comportamiento del niño (Narad *et al.*, 2015) y al menos una parte de las conductas descritas por los síntomas de TDAH podría manifestarse de forma diferente en los contextos familiar y escolar (Gomez, 2007).

El segundo objetivo específico fue investigar si las tres presentaciones del TDAH propuestas por el DSM-5 pueden ser empíricamente validadas. Un modelo factorial mixto que apoye la clasificación propuesta por el DSM debería consistir en dos factores (DA y HI) y al menos cuatro clases que muestren diferentes patrones de síntomas (clase no afectada, clase combinada con alta probabilidad de adhesión a todos los síntomas -reflejando la presentación combinada-, clase predominantemente hiperactiva y clase predominantemente inatenta). Por el contrario, un modelo acorde con la hipótesis de que el TDAH se organiza como un continuo general presente en toda la población, debería dar lugar a clases diferenciadas por la gravedad con la que presentan los síntomas y no por patrones de respuesta cualitativamente distintos.

Método

Participantes

La muestra consistió en 871 niños (465 varones y 406 mujeres) de edades de entre cinco y 14 años ($M= 8,9$; $DT= 2,51$). La muestra fue obtenida de forma incidental en ocho centros escolares de Castilla y León que voluntariamente colaboraron en esta investigación. No se establecieron criterios de exclusión de participantes.

Instrumento

Utilizamos un cuestionario construido a partir de los 18 síntomas del TDAH propuestos por la versión española del DSM-5 (APA, 2014). La justificación del uso de la mencionada escala queda refrendada por el hecho de que es común, a la hora de intentar evaluar la sintomatología del TDAH que propone el DSM, trasladar tales síntomas a escalas similares a la aquí utilizada (Arias *et al.*, 2016b; Amador, Forns, Guàrdia y Perú, 2006; Barkley y Murphy, 2006; Gomez, 2007). Empleamos un procedimiento similar al propuesto por Arias *et al.* (2016a) para la traslación a escalas de clasificación de los síntomas TDAH del DSM-IV-TR. La formulación de los ítems coincidió de forma literal con la versión española del DSM-5, con la salvedad de que se eliminó la expresión adverbial “con frecuencia”, con la que comienza la redacción de todos los síntomas, dado que el uso de tal expresión anularía de facto la escala de frecuencia. Así, por ejemplo, el ítem tres de la escala de déficit de atención fue “*Parece no escuchar cuando se le habla directamente (por ejemplo, parece tener la mente en otras cosas, incluso en ausencia de cualquier distracción aparente)*” y el ítem dos de la escala de hiperactividad “*Se levanta en situaciones en que se espera que permanezca sentado (por ejemplo, se levanta en clase, o en situaciones que requieren mantenerse en su lugar)*”. También se eliminaron las referencias que pudiesen resultar inadecuadas a la evaluación de niños de entre cinco y 14 años (referencias al ámbito laboral en los síntomas tres y seis de déficit de atención y el síntoma dos de hiperactividad, y términos específicos solo apropiados a la evaluación del TDAH en jóvenes y adultos). Por último, cada ítem debía ser valorado en una escala de respuesta de frecuencia de cinco puntos (1= casi nunca, 2= algunas veces, 3= bastantes veces y 4= muchas veces y 5= casi siempre), de acuerdo al comportamiento habitual del niño en los últimos seis meses.

Procedimiento

Los niños fueron evaluados por sus profesores mediante una plataforma *online* habilitada en la Universidad de Valladolid. Previamente a la toma de datos, los investigadores visitaron los colegios y explicaron detalladamente los objetivos del estudio y las instrucciones de respuesta. Así mismo, se habilitó un teléfono y una dirección de correo electrónico a fin de que los respondientes pudieran resolver dudas directamente con el investigador responsable. Dadas las

características de la plataforma de respuesta, no se observaron casos con datos perdidos. Un total de 53 profesores aceptaron participar en el estudio. Cada profesor calificó a los estudiantes identificados con un número impar en listas alfabéticas creadas a partir del primer apellido. El número de niños evaluados por cada profesor osciló entre 13 y 18 ($M= 16,04$; $DT= 0,83$). La tasa de retorno de la información fue del 100%. Las escalas se aplicaron seis meses después del inicio del año escolar, a fin de asegurarse de que los profesores estuvieran suficientemente familiarizados con el comportamiento de los niños. Las evaluaciones fueron completamente anónimas y se contó con consentimiento informado. Todos los procedimientos realizados en este estudio estuvieron de acuerdo con el Código de Prácticas de Investigación de la Universidad de Valladolid (España), así como con la Declaración de Helsinki de 1964 y sus modificaciones posteriores, y normas éticas comparables.

Análisis de datos

AJUSTE DEL MODELO DE MEDIDA. Como paso previo a los análisis principales, evaluamos la calidad del instrumento de medida. Para ello, comparamos el ajuste de tres modelos factoriales confirmatorios: un modelo unidimensional, un modelo de dos factores correlacionados (déficit de atención, con nueve ítems, e hiperactividad-impulsividad, con nueve ítems) y un modelo de tres factores correlacionados (déficit de atención, con nueve ítems, hiperactividad, con seis ítems, e impulsividad, con tres ítems). Dada la investigación previa acerca de la estructura factorial de los síntomas del TDAH propuestos por el DSM, es esperable que el modelo de dos factores correlacionados adquiera mejor ajuste e interpretabilidad (Willcutt *et al.*, 2012).

Evaluamos el ajuste de los modelos mediante el índice de ajuste comparativo (CFI), el índice de Tucker-Lewis (TLI) y el error cuadrático medio de aproximación (RMSEA). En el caso de CFI y TLI, valores superiores a 0,90 y 0,95 sugieren ajuste aceptable y bueno, respectivamente (Hu y Bentler, 1999; Marsh y Hau, 1996). En el caso de RMSEA, valores inferiores a 0,08 y 0,05 sugieren ajuste aceptable y bueno, respectivamente (Browne y Cudeck, 1992). Estimamos los modelos a partir de la matriz de correlaciones policóricas mediante mínimos cuadrados ponderados con media y varianza ajustada (WLSMV), dada la naturaleza ordinal de los datos de entrada (Beauducel y Herzberg, 2006). La posible no independencia entre las observaciones hechas por el mismo profesor fue tenida en cuenta mediante la inclusión en el modelo de un componente multinivel.

Por último, evaluamos la calidad de la medida mediante la inspección de la magnitud de las saturaciones factoriales estandarizadas (cuya varianza media extractada debería ser superior a 0,50 para garantizar suficiente validez convergente de los ítems) y estimamos la fiabilidad de las puntuaciones latentes mediante el cálculo de omega de McDonald (valores próximos a 1 sugieren que las puntuaciones latentes están siendo estimadas con alta precisión; Zinbarg, Revelle, Yovel y Li, 2005).

AJUSTE DE LOS MODELOS FACTORIALES, DE CLASES LATENTES Y FACTORIALES MIXTOS. El plan para la especificación de los modelos y el análisis de los datos se diseñó de acuerdo con las recomendaciones de Clark *et al.* (2013). Se estimaron tres conjuntos de modelos (factorial confirmatorio, de clases latentes y factorial mixto). En el caso de los modelos factoriales confirmatorios, se estimó el ajuste de un modelo unidimensional que hipotetiza la existencia de una única variable continua que explica las respuestas a los 18 síntomas y un modelo de dos factores correlacionados (déficit de atención e hiperactividad/impulsividad, ambos medidos por nueve síntomas). En el caso de los modelos de clases latentes, comenzamos estimando el modelo más parsimonioso (dos clases). En sucesivas iteraciones añadimos una clase en cada modelo, hasta llegar a un modelo con k clases que no mostrase mejor ajuste que el modelo con $k-1$ clases. Por último, diseñamos los modelos factoriales mixtos conforme al ajuste obtenido en los modelos factoriales y del modelo de clases latentes con mayor número de clases. Dicha combinación dio lugar a la estimación de dos modelos factoriales, cuatro modelos de clases latentes y cuatro modelos factoriales mixtos.

Para evaluar el ajuste relativo de los modelos empleamos el criterio de Información Bayesiano (BIC), para el que valores menores sugieren mejor ajuste. Utilizamos también el test de Vuong-Lo-Mendell-Rubin (LMR-LRT; Lo, Mendell y Rubin, 2001), que contrasta la hipótesis de que la solución con k clases es mejor que la solución con $k-1$ clases. Valores significativos del test ($p < 0,05$) sugieren que la solución con k clases explica mejor los datos que la solución inmediatamente más parsimoniosa.

Por otra parte, investigamos indicadores adicionales para evaluar la interpretabilidad de los modelos con mejor ajuste. Inspeccionamos la entropía (con un valor entre cero y uno, que indica el grado en que las clasificaciones de sujetos realizadas por el modelo son precisas), la proporción de sujetos capturados por cada clase (p. ej., clases con proporciones extremadamente bajas posiblemente no sean interpretables desde un punto de vista substantivo) y los perfiles de respuesta asociados a cada clase (p. ej., clases con patrones de respuesta muy próximos podrían no ser interpretables como grupos distintos). Adicionalmente, dado el posible impacto de las variables sexo y edad en la manifestación de los síntomas del TDAH (Willcutt *et al.*, 2012), las añadimos como covariables al modelo, al objeto de evitar posibles efectos confundentes en la asignación de los sujetos a las clases. Todos los análisis se realizaron con Mplus v7.0. (Muthén y Muthén, 2015).

Resultados

Ajuste del modelo de medida

La tabla 1 muestra el ajuste de los tres modelos de medida. El modelo unidimensional adquirió un ajuste inaceptable (CFI= 0,87; TLI= 0,85; RMSEA= 0,099). El modelo de dos dimensiones mejoró el ajuste hasta valores aceptables (CFI= 0,94; TLI= 0,93; RMSEA= 0,071), ajuste que no fue sustancialmente mejor en el modelo de tres dimensiones (CFI= 0,94; TLI= 0,93; RMSEA= 0,071). Dados estos resultados, decidimos conservar el modelo de dos factores, en cuanto que la

pérdida de parsimonia del modelo de tres (con dos grados de libertad menos) no condujo a una mejora relevante en el ajuste.

Las saturaciones estandarizadas del factor de déficit de atención estuvieron comprendidas entre 0,86 y 0,74 ($M= 0,80$; varianza media extractada de 0,64), con un omega de McDonald de 0,94. Las saturaciones del factor de hiperactividad/impulsividad estuvieron comprendidas entre 0,87 y 0,56 ($M= 0,74$, varianza media extractada de 0,59), con un omega de McDonald de 0,92.

Tabla 1
Ajuste de los modelos de medida

Modelo	CFI	TLI	RMSEA	<i>gl</i>	PL
Unidimensional	0,870	0,852	0,099	135	90
Dos factores	0,946	0,938	0,071	134	91
Tres factores	0,947	0,939	0,071	132	93

Nota: CFI= índice de ajuste comparativo; TLI= índice de Tucker Lewis; RMSEA= error cuadrático medio de aproximación; *gl*= grados de libertad; PL= parámetros libres.

Ajuste de los modelos factoriales, de clases latentes y factoriales mixtos

La tabla 2 muestra los resultados de los análisis factoriales confirmatorios (modelos M1 y M2), análisis de clases latentes (modelos M3 a M6) y análisis factoriales mixtos (modelos M7 a M10).

Respecto al análisis factorial, la solución unidimensional demostró peor ajuste que la solución de dos factores correlacionados. En lo que se refiere a las clases latentes, el BIC sugirió la retención de cinco clases, mientras que el test LMR sugirió una solución óptima de tres clases. Decidimos en consecuencia establecer el número máximo de clases en cinco para diseñar los contrastes en los análisis factoriales mixtos.

De acuerdo con los valores de BIC, el análisis factorial mixto produjo como solución óptima cuatro clases con dos factores (M9), y de tres clases y dos factores de acuerdo con el test LMR (M8). La entropía fue razonablemente elevada en ambos casos, sugiriendo que el modelo clasificó a los sujetos con suficiente precisión. Dado que, en caso de incongruencia entre índices de ajuste, el BIC tiende a ser un indicador más preciso que el test LMR (Clark *et al.*, 2013), decidimos retener el modelo de cuatro clases y dos factores.

La figura 1 muestra la probabilidad de que un niño sea clasificado en la categoría cuarta ("muchas veces") en cada uno de los 18 síntomas y según la clase de pertenencia. La clase minoritaria (6%) presentó probabilidades superiores al 50% en todos los síntomas, que oscilaron desde un 92% en el síntoma 10 ("Se mueve en exceso") a un 51% en el síntoma 13 ("Habla demasiado alto"). La clase 2 (15%) mostró elevaciones moderadas o bajas en todos los síntomas. Por último, las clases 3 (51%) y 4 (28%) obtuvieron bajas probabilidades en todos los síntomas, si bien en el caso de la clase 3 se observaron ligeras elevaciones en los síntomas 10 ("Se mueve en exceso") y 15 ("Habla demasiado"). Como se puede observar, ninguna clase presentó patrones de probabilidad claramente distintos al resto (p. ej., con elevadas probabilidades de cumplir los síntomas de déficit de

atención y bajas probabilidades de cumplir los síntomas de hiperactividad/impulsividad).

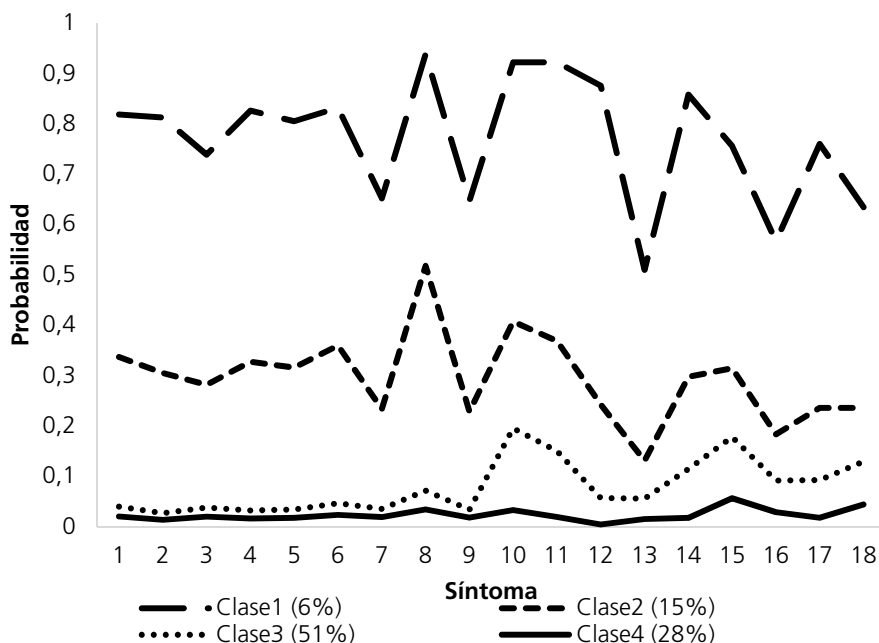
Tabla 2
Resultados de la estimación de los modelos

Análisis	Modelo	Clases	Factores	Par.	BIC	Entropía	LMR $p.$	Dist. clases.
AFC	M1	0	1	90	35690			
	M2	0	2	91	34594			
ACL	M3	2	0	145	37832	0,95	<0,01	73, 26
	M4	3	0	218	36667	0,92	<0,01	41, 40, 17
	M5	4	0	291	36427	0,92	>0,05	37, 27, 18, 16
	M6	5	0	364	36307	0,92	>0,05	38, 24, 18, 11, 6
MFM	M7	2	2	95	34615	0,82	<0,01	80, 20
	M8	3	2	100	34566	0,68	<0,01	49, 40, 11
	M9	4	2	105	34556	0,72	>0,05	51, 28, 15, 6
	M10	5	2	110	34594	0,76	>0,05	50, 37, 5, 3, 2

Nota: AFC= análisis factorial confirmatorio; ACL= análisis de clases latentes; MFM= modelos factoriales mixtos; BIC= criterio de información bayesiano; LMR $p.$ = nivel de significación del test de Vuong-Lo-Mendell-Rubin; Par.= parámetros libres; Dist. clases= proporciones en porcentaje en las que se distribuyen las clases.

Figura 1

Probabilidad por clase de ser calificado en la categoría cuatro o superior en cada síntoma



Nota: Síntomas de 1 a 9: Déficit de atención; síntomas 10 a 18: Hiperactividad-Impulsividad.

Distribución de las clases en la variable latente

Sobre el modelo final (M9, cuatro clases y dos factores) estimamos las puntuaciones latentes obtenidas por los sujetos en cada clase, cuyas distribuciones se muestran en la figura 2. Para facilitar la interpretación de los resultados, las puntuaciones latentes fueron transformadas a una escala con media 100 y desviación típica 15. Como era esperable dadas las características del modelo, las distribuciones estuvieron solapadas en sus colas, si bien las zonas con mayor densidad fueron claramente diferenciables. La clase 1 (6%) presentó la mayor densidad de puntuaciones en zonas altas y muy altas de ambas variables latentes (por encima de dos desviaciones típicas). El resto de las clases presentó diverso grado de variabilidad, pero en todos los casos con puntuaciones en los factores progresivamente menores, hasta la clase 4 (28%) cuya máxima densidad de puntuaciones se observó por debajo de la media. En la misma figura se muestra la distribución de las puntuaciones para el conjunto de todos los sujetos evaluados. Hay claras diferencias de simetría entre la distribución total, que tiende a la asimetría positiva (tal y como es esperable de la aplicación de una escala clínica a una muestra de población general) y las distribuciones de las clases, que tienden a presentar mayor simetría. Para verificar esta impresión visual, estimamos la desviación de la normalidad univariada en cada clase y factor (tabla 3). Aquí se verificó que (a) las distribuciones de la muestra completa presentaron valores de asimetría sustancialmente mayores que las distribuciones de las clases individuales y (b) para las puntuaciones totales se ha de rechazar en ambos factores la hipótesis nula de que las distribuciones univariadas son normales, mientras que no se puede rechazar ($p < 0,05$) en el caso de las distribuciones por clase. Este resultado sugiere que las clases no representaron grupos con perfiles sintomáticos cualitativamente distintos, sino que modelaron la distribución general no normal como una mixtura de distribuciones aproximadamente normales dentro de un mismo continuo latente.

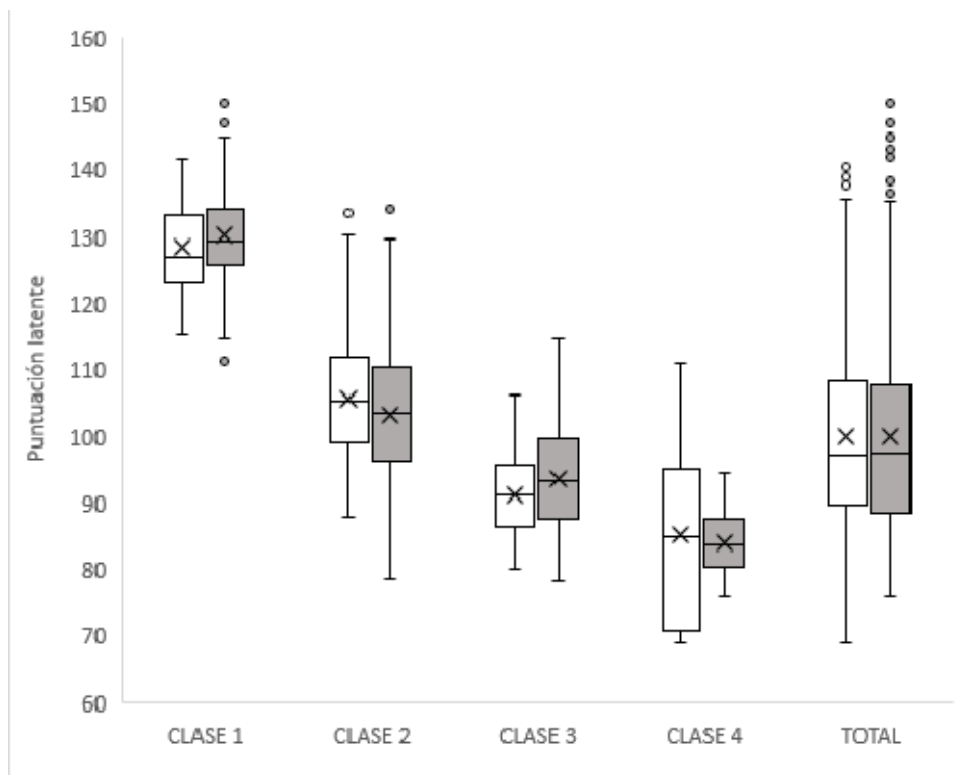
Tabla 3
Normalidad univariada por clase y factor

Variable	Grupo	Asimetría	Curtosis	K-S (sig.)
Déficit de atención	Clase 1	0,225	0,313	0,816
	Clase 2	0,585	-0,818	0,061
	Clase 3	-0,286	-0,059	0,680
	Clase 4	0,312	-0,407	0,188
	Total	0,771	0,264	0
Hiperactividad-impulsividad	Clase 1	-0,075	0,494	0,511
	Clase 2	0,280	0,065	0,416
	Clase 3	0,415	-0,460	0,353
	Clase 4	0,084	-0,612	0,469
	Muestra completa	0,919	0,645	0,001

Nota: K-S (sig.)= nivel de significación de la prueba de normalidad univariada de Kolmogorov-Smirnov.

Figura 2

Distribución de las puntuaciones latentes en cada clase y en la muestra completa



Nota: Las cajas blancas representan la distribución en la variable déficit de atención; las cajas sombreadas representan la distribución en la variable hiperactividad/impulsividad.

Discusión

El principal objetivo de este estudio fue investigar la estructura fenotípica de los síntomas de TDAH del DSM-5 (APA, 2014). Para ello, comparamos modelos puramente continuos, puramente categoriales y mixtos estimados sobre las puntuaciones de profesores recibidas por una amplia muestra de niños pertenecientes a población general. De los modelos, el más plausible fue el compuesto por dos factores (inatención e hiperactividad/impulsividad) y cuatro clases. Las clases difirieron en la gravedad con que presentaron el común de los síntomas TDAH: una clase minoritaria (6%) con elevada gravedad, una clase (15%) con elevaciones moderadas en la mayor parte de los síntomas, una clase (53%) con ligeras elevaciones en algunos síntomas de hiperactividad y una clase (29%) sin elevaciones perceptibles en ningún síntoma.

El modelo final estableció una separación entre una clase minoritaria de alta gravedad y tres clases de gravedad variable. No hay, pues, evidencia clara de que

en esta muestra sea posible identificar subgrupos basados en perfiles sintomáticos cualitativamente distintos. Por el contrario, las clases reflejaron diferencias cuantitativas en la gravedad con que los niños presentaron los síntomas. Para comprobarlo, investigamos si las clases encontradas se correspondían realmente a grupos naturales presentes en la población. La distribución de las variables latentes (inatención e hiperactividad-impulsividad) en toda la muestra fue fuertemente asimétrica (lo que es esperable en la evaluación clínica de población general). Sin embargo, las clases adquirieron distribuciones aproximadamente normales en ambas variables latentes. Los componentes de la mixtura en AFM pueden surgir, aparte de por la presencia de clases naturales con diferentes perfiles de respuesta, como forma de modelar una distribución general fuertemente asimétrica mediante la extracción de dos o más sub-muestras con distribuciones aproximadamente normales en la variable latente (Lubke y Muthén, 2005). En consecuencia, posiblemente en nuestro estudio las clases no fueron reflejo de perfiles conductuales específicos, sino que representaron distribuciones secundarias en un único continuo latente de gravedad, común a toda la muestra. Por esta razón, aunque el modelo con mejor ajuste incluyó tanto dimensiones como clases, se podría argumentar que sería más apropiado un modelo dimensional del TDAH con alguna separación entre niveles más y menos graves de presentación de los síntomas.

En lo que se refiere a la evaluación clínica y el diagnóstico del TDAH, estos resultados no apoyan la clasificación de los niños en subtipos o presentaciones, sino su evaluación en un continuo único donde la presencia o ausencia del TDAH se juzgue en términos de gravedad con que se presenta el espectro de síntomas. Por otra parte, observamos claramente cómo en el grupo potencialmente diagnosticable los síntomas difirieron sustancialmente en la probabilidad de ser observados, desde síntomas con una probabilidad relativamente elevada (p. ej., "se distrae fácilmente") a síntomas con una probabilidad relativamente baja (p.ej., "es olvidadizo en actividades cotidianas"). Esto sugiere la necesidad de refinar el proceso diagnóstico mediante la ponderación de la gravedad del síntoma, e incluso de valorar la posibilidad de definir síntomas marcadores del TDAH, tal y como se ha hecho en otros trastornos como la depresión.

Nuestros resultados difieren respecto de investigaciones anteriores (Gomez *et al.*, 2013, Lubke *et al.*, 2007; Ranby *et al.*, 2012), donde se conservó el modelo con dos factores y dos clases (no afectada y potencialmente afectada). Esta incongruencia de resultados podría en parte deberse a ciertas características propias de nuestro estudio: en primer lugar, hemos analizado las evaluaciones de profesores, mientras que en los estudios citados los análisis se realizaron sobre evaluaciones de padres (Lubke *et al.*, 2007, Gomez *et al.*, 2013) o auto-informes retrospectivos de adultos (Ranby *et al.*, 2012). Dado que el contexto de observación es distinto y usualmente los profesores tienen acceso a muestras de comportamiento diferentes a las disponibles para los padres, es esperable que los resultados difieran en alguna medida. En segundo lugar, dos de los estudios referidos arriba emplearon la escala SWAN (Swanson *et al.*, 2012), mientras que nosotros empleamos una adaptación casi literal de los síntomas propuestos por el DSM-5 (Arias *et al.*, 2016a). Una diferencia fundamental entre ambas medidas es

que el SWAN presenta los 18 síntomas en formato positivo (p. ej., “es organizado” en lugar de “es desorganizado”). En consecuencia, diferencias en la polaridad de los ítems pueden dar lugar a diferencias en los resultados del análisis, ya que las variables medidas con escalas clínicas suelen funcionar como semi-rasgos más que como dimensiones bipolares (Reise y Waller, 2009). En tercer lugar, al contrario que en estudios anteriores, en nuestros análisis hemos incluido edad y sexo como covariables a fin de evitar un posible efecto confundente en la asignación de clases. Esto ha podido contribuir a las diferencias de resultados respecto de estudios donde el efecto de dichas variables no fue controlado. Por otra parte, es necesario tener en cuenta que las diferencias respecto de otros estudios se han observado en las clases de menor gravedad, siendo la proporción estimada de niños potencialmente diagnosticables (6%) similar a la encontrada en estudios anteriores (p. ej., Lubke *et al.* 2007, con un 6.8%), y compatible con la prevalencia del TDAH generalmente estimada en un rango del 3% al 7% de acuerdo a numerosos estudios (APA, 2014).

Nuestros resultados sugieren que el algoritmo diagnóstico propuesto por el DSM, fundamentado en la categorización de sujetos en subtipos o presentaciones mediante conteo no ponderado de síntomas cumplidos (APA, 2014), quizá no sea el procedimiento de evaluación más apropiado. Por el contrario, el trastorno podría ser considerado como un espectro de síntomas de inatención e hiperactividad/impulsividad, donde la presencia o ausencia del trastorno sea constatada por la gravedad general con que un sujeto presenta el común de los síntomas de TDAH. La inserción de un criterio diagnóstico fundado en la gravedad de los síntomas requeriría un importante esfuerzo de investigación que diera el respaldo empírico necesario para establecer normas de decisión firmes. Sin embargo, tendría ventajas importantes: Por ejemplo, ya no se usaría un punto de corte (seis síntomas) universal pero inflexible y en buena parte arbitrario, sino que el punto de corte para considerar a un niño potencialmente diagnosticable podría adaptarse a aspectos que se sabe influyen en la manifestación conductual del TDAH, tales como el estadio del desarrollo o el contexto social y cultural. En términos de tratamiento, pasar de un diagnóstico puramente categorial (presencia o ausencia de TDAH) a uno basado en la evaluación de la gravedad facilitaría el diseño de intervenciones basadas en la evidencia y orientadas a la disminución de la gravedad de las manifestaciones del trastorno (Lubke, Hudziak, Derks, van Bijnsterveldt y Boomsma, 2009).

Los resultados de este estudio pueden ser mejor interpretados teniendo en cuenta sus limitaciones principales. En primer lugar, utilizamos únicamente evaluaciones de profesores. Dado que padres y profesores aportan información válida sobre aspectos distintos de la conducta del niño, comprobar en futuros estudios hasta qué punto la asignación a las clases depende del tipo de informante sin duda ayudaría a entender mejor la estructura mixta del TDAH. En segundo lugar, la evaluación de los síntomas se realizó mediante escalas de clasificación. Para verificar la ausencia de un efecto de método relevante, sería necesario comprobar que los resultados se replican con información obtenida de otras fuentes, especialmente la entrevista clínica y la observación directa en contextos naturales. En tercer lugar, no se empleó en los análisis datos de niños con

diagnóstico formal de TDAH, por lo que se desconoce si en ese caso se hubieran encontrado perfiles equivalentes a los subtipos del DSM. Sin embargo, resultados anteriores con muestras clínicas (Gomez *et al.*, 2013) sugieren una estructura muy similar a la obtenida en nuestro estudio, donde los subtipos son mejor explicados por diferencias en la gravedad del trastorno que por perfiles sintomáticos específicos.

Referencias

- Amador, J. A., Forns, M., Guàrdia, J. y Però, M. (2006). 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, 23-32.
- Asociación Americana de Psiquiatría (1988). *Manual diagnóstico y estadístico de los trastornos mentales, DSM-III-R* (3ª ed. rev.). Barcelona: Elsevier Masson (Orig. 1987).
- Asociación Americana de Psiquiatría (2014). *Manual diagnóstico y estadístico de los trastornos mentales: DSM-5*. Buenos Aires: Editorial Médica Panamericana (Orig. 2013).
- Arias, V. B., Esnaola, I. y Rodríguez-Medina, J. (2018). Identifying potentially marker symptoms of attention-deficit/hyperactivity disorder. *PeerJ*, 6, e4820.
- Arias, V. B., Ponce, F. P. y Núñez, D. (2016a). Bifactor models of attention-deficit/hyperactivity disorder (ADHD): an evaluation of three necessary but underused psychometric indexes. *Assessment*, 20, 1-13.
- Arias, V. B., Ponce, F. P., Martínez-Molina, A., Arias, B. y Núñez, D. (2016b). General and specific attention-deficit/hyperactivity disorder factors of children 4 to 6 years of age: an exploratory structural equation modeling approach to assessing symptom multidimensionality. *Journal of Abnormal Psychology*, 125, 125.
- Barkley, R. A. y Murphy, K. (2006). *Attention-deficit hyperactivity disorder. A clinical workbook*. New York, NY: Guilford.
- Bauermeister, J. J., Canino, G., Polanczyk, G. y Rohde, L. A. (2010). ADHD across cultures: ¿Is there evidence for a bidimensional organization of symptoms? *Journal of Clinical Child and Adolescent Psychology*, 39, 362-372.
- Beauducel, A. y Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, 13, 186-203.
- Borsboom, D., Rhemtulla, M., Cramer, A. O. J., Van der Maas, H. L. J., Scheffer, M. y Dolan, C. V. (2016). Kinds versus continua: a review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychological medicine*, 46, 1567-1579.
- Browne, M. W. y Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21, 230-258.
- Cardo, E., Servera, M. y 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, 10-14.
- Clark, S. L., Muthén, B., Kaprio, J., D'Onofrio, B. M., Viken, R. y Rose, R. J. (2013). Models and strategies for factor mixture analysis: an example concerning the structure underlying psychological disorders. *Structural Equation Modeling: a multidisciplinary journal*, 20, 681-703.
- Gomez, R. (2007). Australian parent and teacher ratings of the DSM-IV ADHD symptoms: differential symptom functioning and parent-teacher agreement and differences. *Journal of Attention Disorders*, 11, 17-27.

- Gomez, R., Vance, A. y Gomez, R. M. (2013). Maternal ratings of the ADHD symptoms: subtypes versus severity in clinic-referred children and adolescents. *Journal of Attention Disorders*, 20, 414-423.
- Hu, L. y Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Hudziak, J. J., Achenbach, T. M., Althoff, R. R. y Pine, D. S. (2007). A dimensional approach to developmental psychopathology. *International Journal of Methods in Psychiatric Research*, 16, S16-S23.
- Hudziak, J. J., Heath, A. C., Madden, P. F., Reich, W., Bucholz, K. K., Slutske, W., Bierut, L. J., Neuman, R. y Todd, R. D. (1998). Latent class and factor analysis of DSM-IV ADHD: a twin study of female adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, 37, 848-857.
- Leoutsakos, J. S., Zandi, P. P., Bandeen-Roche, K. y Lyketsos, C. G. (2010). Searching for valid psychiatric phenotypes: Discrete latent variable models. *International Journal of Methods in Psychiatric Research*, 19, 63-73.
- Lo, Y., Mendell, N. R. y Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767-778.
- Lubke, G. H., Hudziak, J. J., Derks, E. M., van Bijsterveldt, T. y Boomsma, D. I. (2009). Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 1085-1093.
- Lubke, G. H. y Neale, M. C. (2008). Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research*, 41, 499-532.
- Lubke, G.H. y Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, 10, 21-39.
- Lubke, G. H., Muthén, B., Moilanen, I. K., McGough, J. J., Loo, S. K., Swanson, J., May, Y., Taanila, A., Tuula, H., Marjo-Riitta, J. y Smalley, S. L. (2007). Subtypes versus severity differences in attention deficit/hyperactivity disorder in the Northern Finnish Birth Cohort. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46, 1584-1593.
- Marsh, H. W. y Hau, K. (1996). Assessing goodness of fit: is parsimony always desirable? *Journal of Experimental Education*, 64, 364-390.
- McCutcheon, A. L. (1987). *Latent Class Analysis (Quantitative Applications in the Social Sciences)*. Londres: Sage.
- Muthén, B. y Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463-469.
- Muthén, L. K. y Muthén, B. (2015). *Mplus statistical software version 7.31*. Los Ángeles, CA: autor.
- Narad, M. E., Garner, A. A., Peugh, J. L., Tamm, L., Antonini, T. N., Kingery, K. M. y Epstein, J. N. (2015). Parent-teacher agreement on ADHD symptoms across development. *Psychological assessment*, 27, 239.
- Nigg, J. T., Tannock, R. y Rohde, L. A. (2010). What is to be the fate of ADHD subtypes? An introduction to the special section on research on the ADHD subtypes and implications for the DSM-V. *Journal of Clinical and Adolescent Psychology*, 39, 723-725.
- Pickles, A. y Angold, A. (2003). Natural categories or fundamental dimensions: on carving nature at the joints and the rearticulation of psychopathology. *Development and Psychopathology*, 15, 529-551.
- Polanczyk, G., de Lima, M., Lessa, B., Biederman, J. y Rohde, L. A. (2007). The worldwide prevalence of TDAH: a systematic review and metaregression analysis. *The American Journal of Psychiatry*, 164, 942-948.

- Ranby, K. W., Boynton, M. H., Kollins, S. H., McClernon, F. J., Yang, C. M. y Fuemmeler, B. F. (2012). Understanding the phenotypic structure of adult retrospective ADHD symptoms during childhood in the United States. *Journal of Clinical Child and Adolescent Psychology, 41*, 261-274.
- Reise, S. P. y Waller, N. G. (2009). Item response theory and clinical measurement. *Annual review of clinical psychology, 5*, 27-48.
- Swanson, J. M., Schuck, S., Porter, M. M., Carlson, C., Hartman, C. A., Sergeant, J. A. y Wigal, T. (2012). Categorical and dimensional definitions and evaluations of symptoms of ADHD: history of the SNAP and the SWAN rating scales. *The International Journal of Educational and Psychological Assessment, 10*, 51.
- Willcutt, E. G., Nigg, J. T., Pennington, B. F., Solanto, M. V., Rohde, L. A., Tannock, R. y Lahey, B. B. (2012). Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes. *Journal of abnormal psychology, 121*, 991-1010.
- World Health Organization (1991). *International classification of diseases (ICD-10)*. Ginebra: Autor.
- Yung, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62*, 297-330.
- Zinbarg, R. E., Revelle, W., Yovel, I. y Li, W. (2005). Cronbach's alpha, Revelle's beta, and McDonald's omega: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 123-133.

RECIBIDO: 12 de enero de 2018

ACEPTADO: 10 de marzo de 2018