Revista de Psicología Clínica con Niños y Adolescentes
Vol. 3. Nº. 2 - Julio 2016 - pp 23-34

Copyright© 2016 RPCNA
www.revistapcna.com - ISSN 2340-8340

Revista de Psicología Clínica con Niños y Adolescentes

SENA, Sistema de Evaluación de Niños y Adolescentes: proceso de desarrollo y evidencias de fiabilidad y validez

Fernando Sánchez-Sánchez[1], Irene Fernández-Pinto[1], Pablo Santamaría[1], Miguel A. Carrasco[2], & Victoria del Barrio[2]
[1]*TEA Ediciones, España*
[2]*Universidad Nacional de Educación a Distancia, UNED, España*

Resumen

El objetivo del presente trabajo es describir el proceso de desarrollo del SENA, Sistema de evaluación de niños y adolescentes, y analizar las evidencias de fiabilidad y validez que apoyan el uso de sus puntuaciones. El SENA es un sistema de evaluación comprehensivo, multifuente y multidimensional orientado a la identificación de los principales problemas emocionales y conductuales en niños y adolescentes (3 a 18 años). Durante su construcción se ha seguido un riguroso proceso de diseño y desarrollo, ajustándose a las directrices internacionales: evaluación por parte de expertos de distintas áreas y especialidades, realización de estudios piloto (N = 1677) y la consideración de múltiples indicadores para la toma de decisiones sobre cada uno de los ítems y escalas que constituyen la versión definitiva del SENA. Esta se aplicó a una muestra de población general con edades comprendidas entre 3 y 18 años, compuesta por 2550 personas, y a una muestra clínica de 996 participantes. Se examinó la fiabilidad de las escalas e índices mediante el alfa de Cronbach y la correlación test-retest, que fueron superiores a .70 en la inmensa mayoría de ellas. La estructura interna de la prueba se analizó mediante análisis factorial confirmatorio, observándose un adecuado ajuste de los datos al modelo planteado en los índices de la prueba. Las evidencias disponibles resultan favorables y apoyan su uso para los fines para los que fue creado, lo que permite contar con un nuevo instrumento con adecuadas garantías psicométricas para la evaluación de los problemas emocionales y conductuales en niños y adolescentes.

Palabras clave: test, evaluación, psicopatología, niños, adolescentes.

Abstract

SENA, Assessment System for Children and Adolescents: development process and reliability and validity evidences. The aim of this study is to introduce the Children and Adolescents´ Assessment System (SENA) and to analyze its reliability and validity properties. The SENA is a multidimensional, multi-informant, and comprehensive system of assessment designed to identify the children and adolescent´s behavioral and emotional problems from 3 to 18 years old. The design of the final version of this system was created in accordance to the international guidelines consisting of a highly detailed process: analyses of experts in different disciplines, a previous exploratory study (N = 1,677) and a complex decision-making process on items and factors considering multiple statistical and theoretical indicators. The sample was composed of two equivalent groups ranging in age from 3 to 18 years: a general group of 2,550 participants; and a clinical group of 996 participants. The reliability of the scales were analyzed by Cronbach´s alpha and test-retest correlations. Both were higher than .70 in most of the obtained scales. The internal structure was analyzed by confirmatory factor analysis. Results show good fit indexes for the proposed model, and they also display suitable psychometric properties that support the use of this new assessment system for children and adolescents.

Keywords: test, assessment, psychopathology, children, adolescents.

La evaluación de los problemas psicopatológicos infanto-juveniles desempeña un papel crítico en el proceso de intervención temprana sobre los mismos. La identificación precoz de posibles problemas y la puesta en marcha de intervenciones eficaces adecuadamente dirigidas por los datos de la evaluación contribuye a un mejor pronóstico de los posibles trastornos, un menor nivel de afectación y desadaptación, a una mejor respuesta a los tratamientos y a evitar la aparición de complicaciones adicionales o la cronificación de los posibles trastornos (Neil & Christensen, 2009; Calear & Christensen, 2010; Kendall, 2012; del Barrio, 2014).

Existe un gran consenso sobre la conveniencia de, ante los primeros indicios de la posible presencia de problemas, realizar una eva-

Correspondencia:
Fernando Sánchez-Sánchez.
Departamento de I+D+i de TEA Ediciones.
C/ Fray Bernardino Sahagún, 24, 28036, Madrid, España.
E.mail: fernando.sanchez@teaediciones.com

luación inicial comprehensiva y multidimensional que incorpore a la evaluación numerosos constructos y variables relacionadas (Achenbach, McConaughy, & Howell, 1987; del Barrio, 2002a; Hunsley & Mash, 2008). Esta necesidad de realizar un rastreo inicial de un amplio conjunto de problemas radica en la elevada prevalencia de los trastornos psicopatológicos de la infancia y la adolescencia.

Se estima que los trastornos infantiles y adolescentes están presentes en un rango del 10% al 19% de la población infanto-juvenil. Las perturbaciones graves como el autismo, la esquizofrenia o el retraso mental tienen una prevalencia entre el 1% y el 2%, mientras que otras perturbaciones más leves, como los problemas de aprendizaje, pueden llegar al 41%. Los datos en la población española de 6 a 17 años muestran unas prevalencias medias del 5% en problemas de sueño, 5% en problemas de alimentación, 7% en depresión, 12% en problemas de eliminación, 21% en ansiedad, 28% en problemas de conducta, así como una prevalencia del 4% de otro tipo de problemas (p. ej., ansiedad de separación, trastorno obsesivo-compulsivo, estrés postraumático) y un 10% sin diagnóstico definido (Ezpeleta, Guillamón, Granero, de la Osa, Domènech, & Moya, 2007; Ezpeleta, 2014; Polanczyk, Salum, Sugaya, Caye, & Rohde, 2015). Estas cifras medias aproximadas varían en función del sexo, la edad, el nivel de educación y el contexto cultural (Navarro, Moral, Galán, & Beitia, 2012).

Además de las elevadas cifras de prevalencia de los diferentes trastornos es necesario considerar la elevada comorbilidad de los mismos en la etapa infantil y juvenil, tanto de los problemas exteriorizados (TDAH, trastorno de conducta, trastorno negativista desafiante…) como de los problemas interiorizados (depresión, ansiedad, TOC…). En algunos trastornos, como es el trastorno por déficit de atención e hiperactividad (TDAH), diversos estudios clínicos y epidemiológicos indican que entre el 50 % y el 90% de los niños y adolescentes que cumplen criterios para el diagnóstico de este trastorno también los cumplen para al menos otro trastorno psicológico o psiquiátrico (Jensen et al., 2001; Barkley, 2006; Frick & Nigg, 2012). Esta elevada comorbilidad entre los diferentes trastornos impone la necesidad de realizar un rastreo amplio de diversos problemas que permita captar las relaciones entre las diferentes manifestaciones de los mismos.

A esta necesidad de realizar una evaluación amplia y comprehensiva de los posibles problemas se une la de dar una respuesta adecuada a las dificultades y requisitos propios de la evaluación a niños y adolescentes. Existe un sólido consenso en admitir que la metodología de la evaluación en general, y de la evaluación infantil en particular, debe incorporar una aproximación multifuente y multimétodo (Achenbach et al., 1987; Kamphaus & Frick, 1996; Carrasco, 2013; del Barrio, 2002a; de los Reyes & Kazdin, 2005). Es decir, la evaluación debe implementarse en la medida de lo posible mediante diferentes informadores (p. ej., padres, profesores, cuidadores…) y con diferentes técnicas (p. ej., observación, entrevista, cuestionarios…). Además, es necesario adoptar un enfoque evolutivo que tenga en cuenta las características de los evaluados de diferente edad (De Pauw & Mervielde, 2010). A su vez, es necesario ir más allá de la mera presencia de sintomatología y tomar en consideración otras variables relacionadas con las características personales de la persona evaluada (sus vulnerabilidades y sus recursos) y con los diferentes contextos en los que se manifiesta el problema. Desde esta perspectiva, el enfoque multidimensional se evidencia como el más adecuado para evaluar la amplia gama de variables implicadas en los trastornos psicopatológicos y dar cuenta adecuadamente de su inherente comorbilidad.

Además, es necesario concebir la evaluación psicológica como un proceso dinámico de toma de decisiones y contraste de hipótesis (Carrasco, 2013; Fernández-Ballesteros et al., 2003) que comienza por la utilización de instrumentos de amplio espectro, diseñados para realizar un rastreo más general de las variables relevantes para, posteriormente, ir centrando la atención en aquellas áreas donde se localicen las dificultades concretas, utilizando para ello instrumentos más específicos (ya sean tests focalizados en variables concretas, entrevistas diagnósticas, observación natural, pruebas de ejecución, etc.). En este marco, el uso de los cuestionarios o escalas de valoración conductual aparece como una de las estrategias más eficaces para realizar una evaluación inicial amplia de los posibles problemas, ya que permiten obtener información de una gran variedad de problemas y variables en un lapso de tiempo muy breve y con poco coste económico (Achenbach & Rescorla, 2000, 2001; Reynolds, 2010).

No obstante, disponer de la instrumentación adecuada es una condición indispensable para cumplir los requisitos necesarios para una evaluación científica. Los avances que se producen tanto en el campo teórico de la psicopatología como en el de la medición imponen la necesidad de realizar una constante renovación y actualización de los instrumentos de evaluación para incorporar estos nuevos conocimientos. Disponer de pruebas de evaluación adecuadamente construidas y validadas es un primer paso para lograr una evaluación de calidad que ofrezca las garantías necesarias.

Existen numerosos tests de amplio espectro que son frecuentemente utilizados en el ámbito internacional, como el *Sistema de Evaluación de Achenbach* (ASEBA; Achenbach & Rescorla, 2000, 2001), el CBRS (*Comprehensive Behavior Rating Scales*; Conners, 2008), el BASC-2 (*Behavior Assessment System for Children*; Reynolds & Kamphaus, 2004a), el PIC-2 (*Personality Inventory for Children*; Lachar & Gruber, 2001), el PIY (*Personality Inventory for Youth*; Lachar & Gruber, 1995) o el SBS (*Student Behavior Survey*; Lachar, Wingenfeld, Kline, & Gruber, 2000), por citar solo algunos. En España solo se dispone de adaptaciones de algunas de estas pruebas, entre las que destacan la del BASC o la del ASEBA. En el caso del BASC se trata una adaptación de la versión original de 1992 (Reynolds & Kamphaus, 2004b) y ofrece unos baremos amplios a nivel nacional recogidos entre 2001 y 2003. En el caso del ASEBA, solo se disponen de baremos españoles para algunos de los instrumentos parciales que componen la prueba y, más en concreto, los correspondientes al CBCL (*Child Behavior Checklist*) deben considerarse provisionales, según los autores de la adaptación española (Unitat d'Epidemiologia i de Diagnòstic en Psicopatologia del Desenvolupament, 2013).

En este contexto, se consideró necesario disponer de una prueba de nueva generación para la evaluación comprehensiva de los principales problemas psicopatológicos de los niños y adolescentes que incluyera no solo los principales trastornos recogidos en el DSM-5 (APA, 2013) sino que también considerara otras variables de gran relevancia clínica, como los factores de vulnerabilidad y de recursos, que estuviera desarrollada íntegramente en español y atendiendo a las especificidades culturales de nuestro entorno y que dispusiera de unos baremos amplios, representativos y actuales. Esto motivó el diseño y desarrollo del SENA, *Sistema de evaluación de niños y adolescentes*.

Los ejes principales que guiaron el desarrollo del SENA fueron los siguientes: 1) la prueba debía representar adecuadamente una amplia variedad de problemas psicológicos propios de niños y adolescentes (enfoque multidimensional); 2) permitir una evaluación adaptada a la edad, contemplando aquellas manifestaciones o dificultades más habituales en cada etapa (enfoque evolutivo);

3) permitir combinar las valoraciones de las principales fuentes de información sobre la conducta de los niños y adolescentes en los diferentes contextos en los que se desenvuelve, esto es, familia, escuela y autoinforme (enfoque multifuente); 4) incorporar a la evaluación, más allá de los problemas, otras variables complementarias, útiles para el planteamiento de pautas de intervención y la prevención de problemas; 5) disponer de evidencias sólidas de fiabilidad y validez para los usos para los que ha sido concebida la prueba; y 6) ofrecer unos baremos representativos, robustos y actuales que permitieran la interpretación de las puntuaciones con confianza.

El objetivo del presente trabajo es describir brevemente el procedimiento seguido para el desarrollo del SENA y analizar las evidencias disponibles sobre la fiabilidad y la validez de sus puntuaciones. En concreto, se analizará la fiabilidad de cada una de sus escalas e índices en términos de su consistencia interna y de su estabilidad temporal (test-retest). En segundo lugar se examinarán diferentes aproximaciones al concepto de validez interna atendiendo a la estructura de las escalas e índices. Las evidencias relacionadas con la validez externa disponibles (capacidad para diferenciar diferentes grupos clínicos y el patrón de correlaciones con otras escalas que evalúan constructos similares) no se incluyen en este trabajo debido a su extensión, pudiendo consultarse en Fernández-Pinto, Santamaría, Sánchez-Sánchez, Carrasco & del Barrio (2015a).

Método

Participantes

Para los análisis realizados se emplearon dos tipos de muestras: la muestra de tipificación y la muestra clínica. La muestra de tipificación estuvo formada por un total de 2550 personas con edades comprendidas entre los 3 y los 18 años (Tabla 1). Los criterios de selección de esta muestra incluían la ausencia de diagnósticos de trastornos psicopatológicos o del desarrollo, por lo que se considera una muestra de población general o comunitaria, en comparación con la muestra clínica que se describe más adelante. Estos participantes fueron evaluados con los diferentes cuestionarios del SENA aplicables en función de su edad, de forma que se disponía de, al menos, dos fuentes de información en los participantes del nivel Infantil (3-6 años; ejemplares para la familia y para la escuela) y de tres fuentes en los niveles Primaria y Secundaria (6 a 12 años y 12 a 18 años, respectivamente, con ejemplares para la familia, para la escuela y de autoinforme). De este modo se obtuvieron las respuestas de un total de 6,734 personas a los distintos ejemplares del SENA. Se utilizó un muestreo estratificado por cuotas para la selección de la muestra, utilizando como variables de estratificación la edad, el sexo y la región geográfica y manteniendo las proporciones recogidas en el padrón municipal de España (INE, 2014). Además, se registraron

Tabla 1. Distribución de la muestra de tipificación en cada cuestionario en función de la edad y el sexo.

	Cuestionario								
	Infantil		Primaria				Secundaria		
Edad (años)	Familia n (%)*	Escuela n (%)*	Autoinforme 6-8 años n (%)*	Autoinforme 8-12 años n (%)*	Familia n (%)*	Escuela n (%)*	Autoinforme n (%)*	Familia n (%)*	Escuela n (%)*
3	113 (49.6%)	113 (49.6%)							
4	138 (50.0%)	140 (50%)							
5	158 (50.6%)	159 (50.9%)							
6	31 (64.5%)	31 (64.5%)	103 (48.5%)		106 (49.1%)	106 (49.1%)			
7			154 (52.6%)		166 (51.8%)	155 (52.9%)			
8			61 (47.5%)	102 (55.9%)	165 (53.3%)	155 (52.3%)			
9				150 (44.7%)	149 (44.3%)	147 (44.9%)			
10				140 (46.4%)	142 (46.5%)	139 (47.5%)			
11				155 (52.3%)	140 (53.6%)	140 (53.6%)			
12				93 (50.5%)	74 (48.6%)	71 (47.9%)	96 (56.3%)	110 (57.3%)	139 (55.4%)
13							136 (50.0%)	143 (49.7%)	176 (50.0%)
14							156 (48.7%)	163 (47.2%)	194 (49.0%)
15							161 (51.6%)	167 (49.1%)	189 (48.7%)
16							158 (51.3%)	163 (52.1%)	173 (50.9%)
17							149 (49.0%)	160 (48.1%)	160 (48.1%)
18							78 (53.8%)	82 (52.4%)	85 (52.9%)
Total	440 (51.1%)	443 (51.2%)	318 (50.3%)	640 (49.5%)	942 (49.8%)	913 (49.9%)	934 (51.1%)	988 (50.4%)	1116 (50.4%)

** Entre paréntesis el porcentaje de mujeres.*

otras variables relevantes como el entorno de residencia de los participantes (i.e., rural, urbano o intermedio), el tipo de centro educativo al que asistían (i.e., concertado, privado o concertado) y el nivel socioeconómico estimado.

En la Tabla 1 se muestra la distribución de la muestra en función de la edad y el sexo. En cuanto a la región geográfica, se incluyeron participantes procedentes de 172 localidades pertenecientes a 34 provincias de 14 comunidades autónomas diferentes repartidas por todo el territorio español. En líneas generales, las proporciones de participantes de cada edad, sexo y región geográfica se ajustaron satisfactoriamente a las proporciones previstas como representación de la población española de entre 3 y 18 años. Se pueden consultar más detalles sobre el muestreo y la composición de la muestra en (Fernández-Pinto et al., 2015a).

La muestra clínica es una muestra incidental formada por un total de 996 participantes de 3 a 18 años (Tabla 2) que habían recibido un diagnóstico formal (según criterios DSM-5; APA, 2013) de algunos de los trastornos contemplados en el estudio. Estos participantes fueron evaluados mediante los ejemplares correspondientes a cada nivel del SENA y procedían en su mayoría de gabinetes o clínicas privadas. El 31.10% de esta muestra clínica estuvo formada por mujeres, variando ligeramente esta proporción en los diferentes niveles de edad: Infantil, n = 136 (19.90% mujeres); Primaria, n = 549 (32.40% mujeres) y Secundaria, n = 311 (33.80% mujeres).

Instrumentos

El SENA, *Sistema de Evaluación de Niños y Adolescentes* (Fernández-Pinto et al., 2015a, 2015b) está compuesto por un total de 9 cuestionarios dirigidos a tres niveles de edad: Infantil (3-6 años), Primaria (6-12 años) y Secundaria (12-18 años). Incluye cuestionarios específicos para recabar información de diferentes informadores atendiendo a los principales contextos en los que se desenvuelve el niño (cuestionarios para familia y para la escuela), además de tres modelos de autoinforme según la edad de la persona evaluada que se aplican a partir de los 6 años.

Cada cuestionario consta de un número de ítems que oscila entre 77 y 188 en el caso de los autoinformes, entre 122 y 141 en los cuestionarios para la escuela y entre 129 y 154 en los dirigidos a la familia. En cada uno de ellos los informadores deben valorar la frecuencia de aparición de la conducta descrita mediante una escala de cinco puntos (Nunca o casi nunca a Siempre o casi siempre) salvo en el autoinforme para los 6-8 años, que incluye una escala de tres opciones: Sí, No y A veces.

Los diferentes modelos de cuestionario tienen un enfoque multidimensional y permiten la evaluación de un amplio conjunto de contenidos mediante tres bloques de escalas: escalas de problemas, de vulnerabilidad y de recursos personales. Las escalas de problemas

Tabla 2. Distribución de la muestra clínica en cada cuestionario en función de la edad y el sexo.

	Cuestionario								
	Infantil		Primaria				Secundaria		
Edad (años)	Familia n (%)*	Escuela n (%)*	Autoinforme 6-8 años n (%)*	Autoinforme 8-12 años n (%)*	Familia n (%)*	Escuela n (%)*	Autoinforme n (%)*	Familia n (%)*	Escuela n (%)*
3	44 (13.6%)	42 (14.3%)							
4	41 (19.5%)	39 (17.9%)							
5	41 (24.4%)	40 (25%)							
6	14 (35.7%)	13 (38.5%)	46 (21.7%)		56 (21.4%)	51 (25.5%)			
7			66 (22.7%)		71 (23.9%)	63 (23.8%)			
8			35 (28.6%)	72 (48.6%)	99 (36.4%)	82 (31.7%)			
9				80 (42.5%)	79 (41.8%)	71 (43.7%)			
10				93 (3.1%)	89 (25.8%)	75 (33.3%)			
11				71 (28.2%)	67 (28.4%)	63 (28.6%)			
12				35 (42.9%)	35 (40%)	29 (37.9%)	45 (40.0%)	38 (44.7%)	31 (48.4%)
13							61 (37.7%)	52 (38.5%)	46 (37.0%)
14							66 (30.3%)	60 (31.7%)	56 (32.1%)
15							36 (41.7%)	33 (39.4%)	29 (37.9%)
16							26 (42.3%)	27 (44.4%)	26 (42.3%)
17							26 (57.7%)	23 (60.9%)	17 (58.8%)
18							8 (25.0%)	9 (22.2%)	7 (28.6%)
Total	140 (2.7%)	134 (2.9%)	147 (23,8%)	351 (37,6%)	496 (31%)	434 (32%)	268 (38.8%)	242 (40.1%)	212 (39.6%)

* *Entre paréntesis el porcentaje de mujeres.*

se organizan a su vez en diferentes grupos atendiendo a la manifestación principal de los trastornos que evalúan. Así, en el bloque de problemas interiorizados se incluyen las escalas Depresión (DEP), Ansiedad (ANS), Ansiedad Social (ASC), Quejas somáticas (SOM), Sintomatología postraumática (PST) y Obsesión-compulsión (OBS). En el grupo de problemas exteriorizados se incluyen las escalas de Inatención (ATE), Hiperactividad-impulsividad (HIP), Problemas de control de la ira (IRA), Agresión (AGR), Conducta desafiante DES) y Conducta antisocial (ANT). En un tercer bloque se incluyen varias escalas que permiten evaluar otro tipo de problemas específicos, como son Problemas de la conducta alimentaria (ALI), Retraso en el desarrollo (RET), Problemas de aprendizaje (APR), Esquizotipia (ESQ), Consumo de sustancias (SUS) y Comportamiento inusual (INU). El último bloque está integrado por las escalas de problemas contextuales: Problemas familiares (FAM), Problemas con la escuela (ESC) y Problemas con los compañeros (COM).

Además de las escalas de problemas, el SENA incluye otras para evaluar algunas variables relevantes para la comprensión del caso y para orientar las posibles intervenciones. El bloque de escalas de vulnerabilidades permite evaluar aspectos como los Problemas de regulación emocional (REG), Rigidez (RIG), Aislamiento (AIS), Dificultades de apego (APE) y Búsqueda de sensaciones (BUS). El último bloque está formado por las escalas de recursos personales, que incluye Autoestima (AUT), Integración y competencia social (SOC), Inteligencia emocional (IEM), Conciencia de los problemas (CNC) y Disposición al estudio (EST).

A partir de la combinación de las escalas anteriores, el SENA permite obtener varios índices: Índice global de problemas, Índice de problemas emocionales, Índice de problemas conductuales, Índice de problemas en las funciones ejecutivas, índice de problemas contextuales e Índice de recursos.

Procedimiento

El primer paso para la construcción del test fue detallar minuciosamente los objetivos que se pretendía conseguir con su utilización y las características fundamentales que debía tener (Abad, Olea, Ponsoda, & García et al., 2011), incluyendo el uso que se hará de las puntuaciones o el tipo de decisiones que se tomarán a partir de ellas (Martínez-Arias, 1995; Martínez-Arias, Hernández-Lloreda, & Hernández-Lloreda, 2006).

A partir de una extensa revisión de la literatura científica sobre la evaluación y la psicopatología en la infancia y la adolescencia se estableció el mapa de los constructos que se debían evaluar y una definición de cada uno de ellos, detallando las diferentes facetas de los mismos que debían incluirse (puede consultarse con detalle la literatura revisada para la creación de la prueba en Fernández-Pinto, 2015). A partir de estos documentos se construyeron las tablas de especificación de los contenidos y de los comportamientos concretos que se debían incluir en los ítems para cubrir las diferentes facetas de cada constructo. Siguiendo estas tablas se creó un banco inicial compuesto por 2553 ítems, debido a la gran cantidad de constructos a evaluar y a la necesidad de redactar ítems idóneos para los distintos rangos de edad e informadores (padres, profesores…). Este banco fue sometido a un proceso sistemático de control de calidad y depuración para seleccionar los más adecuados (Abad et al., 2011), incluyendo tanto análisis cualitativos (p. ej., revisión de expertos) como cuantitativos (análisis de ítems, fiabilidad, validez…) (Downing & Haladyna, 1997).

Siguiendo las recomendaciones de Martínez-Arias et al. (2006), la revisión cualitativa se realizó mediante las valoraciones de varios jueces expertos en diferentes áreas: expertos en los contenidos evaluados, expertos en psicometría y medición y expertos del ámbito aplicado que trabajan directamente con la población objetivo. Además, se consideró necesario incluir una valoración adicional de expertos referida a la perspectiva lingüística y cultural que garantizara la idoneidad de los ítems del SENA para su uso en todos los países hispanohablantes. En total, 72 expertos participaron en las distintas labores de revisión de los ítems de la prueba.

Paralelamente, se realizó un estudio piloto en el que se aplicaron las versiones experimentales de cada modelo de cuadernillo a una muestra total de 1677 casos de población general (cada cuadernillo se aplicó a un mínimo de 89 casos y un máximo de 661) y 237 clínicos. A partir de estas muestras se realizaron diversos análisis estadísticos para comprobar el funcionamiento de los ítems y de las escalas. En primer lugar, se analizó la fiabilidad de la escala (consistencia interna) y la homogeneidad de cada uno de sus ítems (correlación ítem-total corregida), la frecuencia con que se escogían sus opciones de respuesta y si existía un número de omisiones llamativamente superior al de los restantes ítems del cuestionario. A continuación se calculó el estadístico *d* de Cohen (1996), comparando las diferencias de cada ítem, y de la escala en su conjunto, entre las muestras de población general y clínica. También se obtuvieron las correlaciones de cada uno de los ítems con las puntuaciones de las restantes escalas del test para detectar aquellos ítems que no eran representantes claros de su escala y que estaban contribuyendo al solapamiento entre constructos. Por último, se calcularon las intercorrelaciones entre los ítems de cada escala, con el fin de detectar ítems demasiado redundantes entre sí ($r > .80$) y que pudieran elevar artificialmente la consistencia interna de la escala.

Además de los análisis anteriores, se estudió la estructura en torno a la cual se agrupaban los ítems y las escalas. Para ello se calcularon las intercorrelaciones entre las escalas del estudio piloto y se realizaron análisis factoriales exploratorios tanto de cada escala como agrupando escalas de contenidos similares.

Todas estas informaciones cualitativas y cuantitativas se utilizaron conjuntamente para seleccionar los ítems que se incluyeron en la versión de tipificación de cada uno de los cuestionarios, eliminándose entre un 19% y un 32% de los ítems de cada cuadernillo. Las versiones de tipificación de cada cuadernillo fueron aplicadas a las muestras descritas en el apartado de Participantes y, a partir de los resultados, se procedió a realizar una nueva depuración de los ítems del instrumento para configurar la versión definitiva. Esta segunda depuración de los ítems del instrumento se realizó con el objetivo de seleccionar un conjunto de ítems óptimo para cada escala, reduciendo en todo lo posible la longitud de las mismas. Para ello se siguieron los mismos criterios y análisis aplicados con los resultados del estudio piloto (correlación ítem-total corregida ≥ .30; correlación con la propia escala mayor que con cualquier otra escala; saturación ≥ .40 en el factor subyacente a la escala; capacidad discriminativa en las muestras clínicas $d \geq .30$; correlación con otros ítems de la misma escala ≤ .80 y otros criterios cualitativos identificados por los aplicadores que habían realizado las evaluaciones). La selección definitiva de los ítems siempre se supeditó a la adecuada representación de los contenidos recogidos en la tabla de especificación para cada escala y para cada faceta de las mismas. Tras este proceso sistemático de selección de los ítems óptimos, se eliminaron entre un 19% y un 31% de los ítems de cada cuadernillo.

Una vez configuradas las escalas definitivas de cada cuestionario se realizaron varios análisis para establecer la fiabilidad de las puntuaciones (i.e., consistencia interna, estabilidad temporal y concordancia

entre informadores) y para recabar diferentes evidencias de validez, tanto internas (i.e., análisis factorial exploratorio y confirmatorio) como externas (i.e., correlaciones con otras pruebas específicas que evalúan uno o varios de los constructos incluidos en el SENA; diferencias de puntuaciones en grupos clínicos específicos y relación con otras variables de interés). En los siguientes apartados se exponen los principales resultados de los análisis de fiabilidad y de la estructura de la prueba.

Análisis estadísticos

En primer lugar se procedió a calcular, como estimador de la consistencia interna, el coeficiente alfa de Cronbach (1951) de cada una de las escalas, tanto en la muestra de tipificación como en la muestra clínica. Esto se hizo en cada uno de los grupos de edad considerados para los baremos de la prueba y empleando el módulo correspondiente del programa SPSS V.19 (IBM Corp., 2010). El cálculo de los coeficientes de fiabilidad promedio de cada escala (incluyendo las diferentes edades) se realizó mediante la transformación *Z* de Fisher. El cálculo de los coeficientes alfa de los índices se realizó mediante la fórmula para la fiabilidad de un test compuesto (Martínez-Arias, 1995).

Para analizar la estabilidad temporal de las puntuaciones se realizó un diseño de recogida de datos específico en el que se aplicó la prueba en varios momentos temporales (test-retest) al mismo grupo de personas (AERA, APA, & NCME, 2014; Anastasi & Urbina, 1997). Estas aplicaciones se realizaron sobre una muestra seleccionada aleatoriamente de la muestra total de tipificación (n = 1151). El intervalo transcurrido entre ambas aplicaciones fue de media 29 días (*DT* = 11.2). Posteriormente se procedió a calcular la correlación producto-momento de Pearson entre las puntuaciones típicas de las escalas (escala T, con media 50 y *desviación típica* 10) obtenidas en las dos aplicaciones de la prueba y se calculó el estadístico *d* de Cohen (1996) como estimador del tamaño del efecto de las diferencias entre las medias de ambas aplicaciones. Las correlaciones entre las puntuaciones fueron corregidas atendiendo a la restricción de rango (Martínez-Arias et al., 2006).

El análisis de la estructura interna del tests se realizó mediante dos tipos de aproximaciones: análisis factoriales confirmatorios (AFCs) y exploratorios (AFEs).

Los análisis factoriales confirmatorios se llevaron a cabo con los 9 cuestionarios a partir de las muestras de tipificación del SENA (muestra de población general). Para realizar estos análisis se utilizó el programa Mplus 4.1 (Muthén & Muthén, 2006). Los modelos factoriales se valoraron atendiendo a la información proporcionada por varios índices de bondad de ajuste como son el índice de ajuste comparativo (CFI; *Comparative Fit Index*; Bentler, 1990) y el índice de Tucker-Lewis (TLI; *Tucker-Lewis Index*; Tucker & Lewis, 1973). El rango de valores que pueden adoptar el CFI y el TLI se sitúa entre 0 y 1, donde 1 indica un ajuste perfecto al modelo. Hu y Bentler (1999) recomiendan considerar valores superiores a .90 en estos índices como indicativos de un buen ajuste. También se consideraron otros indicadores de ajuste como la raíz media cuadrática del error de aproximación (RMSEA; Steiger, 1990) y el residuo cuadrático medio estandarizado (SRMR; Bentler, 1995). Los valores de RMSEA iguales o inferiores a .05 indican un buen ajuste al modelo mientras que valores situados entre .05 y .08 representan un ajuste adecuado con errores de aproximación a la población razonables (Browne & Cudeck, 1993). Hu y Bentler (1999) consideran el SRMR como el índice más sensible a las malas especificaciones del modelo y sugieren un valor umbral de .08 para discriminar modelos caracterizados por un buen ajuste a los datos respecto a modelos peores.

Estos análisis se llevaron a cabo únicamente con las escalas de problemas interiorizados y exteriorizados del SENA, por ser las escalas principales del instrumento y ser comunes en la gran mayoría de los ejemplares. Por cada uno de ellos se analizó el ajuste a varios modelos de complejidad creciente:

- Modelo de un factor: todas las escalas pesan en un único factor general.
- Modelo de dos factores: las escalas se separan en dos factores, Interiorizado y Exteriorizado.
- Modelo de tres factores: las escalas se separan en tres factores, Emocional (que coincide con el interiorizado del modelo anterior), Función ejecutiva (incluye las escalas Problemas de atención e Hiperactividad-impulsividad, que junto con las escalas de vulnerabilidades conformarán el Índice de problemas en las funciones ejecutivas del SENA) y Conductual (incluye las escalas de problemas exteriorizados con la excepción de las incluidas en el factor anterior).
- Modelo de tres factores modificado: se trata del mismo modelo anterior pero se introducen pequeñas variaciones sugeridas por los índices de modificación. Se fue muy conservador a la hora de introducir modificaciones sugeridas por estos índices, ya que estos cambios realizados *ad hoc* pueden producir un fenómeno de capitalización del azar si no se replican con otros datos (Abad et al., 2011).

Tabla 3. Consistencia interna (alfa de Cronbach) de las escalas e índices del SENA en la muestra normal y clínica.

Escala / Índice	Cuestionario																	
	Familia						Escuela						Autoinforme					
	Infantil		Primaria		Secundaria		Infantil		Primaria		Secundaria		Primaria 6-8		Primaria 8-12		Secundaria	
	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica
DEP	.84	.81	.86	.89	.90	.91	.89	.90	.91	.91	.91	.90	.81	.79	.84	.85	.91	.92
ANS	.76	.74	.84	.84	.84	.84	.83	.85	.87	.88	.87	.88	.75	.77	.85	.81	.86	.88
ASC	.82	.87	.84	.83	.84	.84	.88	.88	.88	.89	.88	.88	---	---	.79	.78	.85	.82
SOM	.73	.74	.76	.77	.79	.79	.84	.87	.88	.87	.84	.89	---	---	.78	.74	.79	.78
PST	---	---	---	---	---	---	---	---	---	---	---	---	---	---	.77	.82	.82	.85
OBS	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	.70	.71

Escala / Índice	Cuestionario																	
	Familia						Escuela						Autoinforme					
	Infantil		Primaria		Secundaria		Infantil		Primaria		Secundaria		Primaria 6-8		Primaria 8-12		Secundaria	
	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica	Normal	Clínica
ATE	.84	.90	.89	.87	.86	.85	.89	.88	.90	.87	.90	.90	.77	.80	.89	.89	.90	.90
HIP	.82	.84	.88	.88	.87	.90	.92	.91	.91	.91	.92	.94	.82	.80	.83	.87	.85	.86
IRA	.85	.88	.88	.91	.87	.91	.90	.94	.91	.94	.89	.94	.72	.77	.84	.85	.83	.87
AGR	.73	.81	.87	.90	.78	.88	.88	.93	.95	.96	.91	.93	---	.80	.80	.81	.75	.78
DES	.78	.82	.85	.86	.82	.87	.87	.91	.84	.90	.87	.92	---	---	.73	.79	.70	.73
ANT	---	---	---	---	.67	.81	---	---	---	---	.71	.92	---	---	---	---	.70	.82
INU	.68	.82	.81	.84	.76	.84	.68	.86	.85	.90	.82	.87	---	---	---	---	---	---
RET	.80	.91	---	---	---	---	.84	.87	---	---	---	---	---	---	---	---	---	---
APR	---	---	---	---	---	---	---	---	.90	.89	.85	.85	---	---	---	---	---	---
ALI	---	---	---	---	.78	.79	---	---	---	---	---	---	---	---	---	---	.84	.81
SUS	---	---	---	---	.74	.67	---	---	---	---	---	---	---	---	---	---	.79	.90
ESQ	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	.77	.80
FAM	---	---	---	---	---	---	---	---	---	---	---	---	.64	.74	.66	.70	.81	.80
ESC	---	---	---	---	---	---	---	---	---	---	---	---	---	---	.77	.72	.81	.80
COM	---	---	---	---	---	---	---	---	---	---	---	---	.83	.77	.83	.84	.84	.83
REG	.75	.83	.81	.84	.83	.84	.87	.92	.87	.89	.86	.86	---	---	.82	.80	.86	.86
RIG	.68	.72	.75	.79	.80	.81	.77	.79	.80	.80	.85	.85	---	---	---	---	---	.86
AIS	.84	.91	.79	.87	.82	.90	.87	.94	.90	.92	.90	.90	---	---	---	---	---	---
APE	.70	.72	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
BUS	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	.87	---
SOC	.84	.86	.81	.86	.80	.84	.85	.90	.90	.90	.89	.89	---	---	.76	.78	.83	.81
IEM	.81	.84	.78	.80	.82	.83	.85	.87	.86	.85	.83	.83	---	---	---	---	---	---
AUT	---	---	---	.77	---	---	---	---	---	---	---	---	.70	.73	.74	.75	.84	.82
CNC	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	.81	.80
EST	---	---	.72	---	.84	.85	---	---	.81	.79	.85	.85	---	---	---	---		---
EMO	.89	.87	.92	.92	.93	.93	.93	.94	.95	.95	.95	.95	.86	.87	.94	.94	.95	.95
CON	.91	.90	.94	.95	.93	.96	.95	.97	.96	.97	.95	.98	.84	.86	.91	.92	.90	.93
EJE	.91	.89	.93	.94	.94	.94	.94	.94	.95	.95	.95	.95	---	---	---	.94	.94	.94
CTX	---	---	---	---	---	---	---	---	---	---	---	---	.81	.83	.86	.86	.89	.87
REC	.88	.94	.87	.89	.90	.90	.91	.93	.93	.92	.93	.92	---	---	.83	.84	.89	.86
GLO	.94	.94	.96	.96	.96	.96	.96	.96	.97	.97	.97	.97	.93	.93	.97	.97	.96	.97

---: Escala no presente en esa versión del ejemplar del SENA; AGR: Agresión; AIS: Aislamiento; ALI: Problemas de la conducta alimentaria; ANS: Ansiedad; ANT: Conducta antisocial; APE: Dificultades de apego; APR: Problemas de aprendizaje; ASC: Ansiedad social; ATE: Problemas de atención; AUT: Autoestima; BUS: Búsqueda de sensaciones; CNC: Conciencia de los problemas; COM: Problemas con los compañeros; CON: Índice de problemas conductuales; CTX: Índice de problemas contextuales; DEP: Depresión; DES: Conducta desafiante; EJE: Índice de problemas en las funciones ejecutivas; EMO: Índice de problemas emocionales; ESC: Problemas con la escuela; ESQ: Esquizotipia; EST: Disposición al estudio; FAM: Problemas familiares; GLO: Índice global de problemas; HIP: Hiperactividad-impulsividad; IEM: Inteligencia emocional; INU: Comportamiento inusual; IRA: Problemas de control de la ira; OBS: Obsesión-compulsión; PST: Sintomatología postraumática; REC: Índice de recursos personales; REG: Problemas de regulación emocional; RET: Retraso en el desarrollo; RIG: Rigidez; SOC: Integración y competencia social; SOM: Quejas somáticas; SUS: Consumo de sustancias.

Resultados

Los valores del coeficiente alfa obtenidos para las escalas del SENA (Tabla 3) se sitúan por encima de .70 en prácticamente todos los casos y por encima de .80 muy frecuentemente. En el caso de las escalas nucleares del SENA, que pretenden recoger los principales problemas emocionales y de conducta (p. ej., Depresión, Ansiedad, Hiperactividad-impulsividad, etc.), los valores que se obtienen son generalmente superiores a .80 y frecuentemente a .90. Solo algunas escalas concretas mostraron valores inferiores a .70, fundamentalmente debido a la escasa variabilidad que presentaban, lo cual es esperable debido a la gravedad de los problemas que evalúan, muy infrecuentes en la población (p. ej., Comportamiento inusual o Conducta antisocial). La consistencia interna de los índices es superior a la de las escalas y se observan valores superiores a .80 en todos los casos y muy próximos o superiores a .90 en la mayoría de ellos.

En la Tabla 4 se muestran los coeficientes de fiabilidad test-retest de las distintas escalas e índices del SENA. Como puede apreciarse, se observan valores por encima de .80 y .90 en la mayoría de los casos. Las puntuaciones obtenidas en la primera y segunda evaluación son muy similares entre sí, con tamaños del efecto entre pequeños y muy pequeños (todas las $d < .23$ y la mayoría $d < .10$) acorde a la escasa susceptibilidad de la prueba a los efectos de la práctica, el aprendizaje o la maduración que pueden presentar otro tipo de instrumentos (p. ej., de desarrollo o inteligencia).

En la Tabla 5 se presentan los índices de ajuste obtenidos en los modelos de 1, 2, 3 y 3 factores modificados aplicados en cada uno de los cuestionarios del SENA. En todos los cuestionarios se observa un incremento progresivo del grado de ajuste de los distintos modelos propuestos a los datos. El primer modelo, unifactorial, presenta unos niveles de ajuste muy pobres en todos los casos, lo cual indicaría que la complejidad de constructos evaluados en las escalas principales del SENA no se puede atribuir a un único rasgo latente. El segundo modelo, de dos factores, muestra mejoras sustanciales en todos los índices de bondad de ajuste aunque continúa arrojando un ajuste pobre en todos los cuestionarios, si bien en el caso de los autoinformes el ajuste se aproxima más a los umbrales deseables. Parece, por tanto, que esta diferenciación entre lo interiorizado y lo exteriorizado constituye una mejor representación de la información contemplada en el SENA, pero aun así no permite explicar varias fuentes de variabilidad importantes. El modelo de tres factores, hipotetizado teóricamente en el SENA, supone una mejora con respecto a los modelos anteriores y en los cuestionarios de autoinforme llega a presentar un ajuste razonable en la mayoría de los índices de bondad de ajuste considerados. La inclusión del tercer factor relativo a las funciones ejecutivas permite contemplar de manera específica la relación más estrecha existente entre las variables Problemas de atención e Hiperactividad-impulsividad y diferenciarlas de otros problemas de conducta con los que guardan estrecha relación pero con los que covarían en menor grado.

El modelo que presenta un mejor ajuste en todos los cuestionarios es el modelo final de tres factores en el que se incluyeron varias modificaciones. Tal y como se comentó anteriormente, se fue muy conservador en la inclusión de estas modificaciones, que solo se realizaron cuando existía una justificación conceptual sólida para esa modificación y la modificación era sugerida y resultaba oportuna en todos los cuestionarios del SENA. Las modificaciones introducidas fueron: a) permitir la correlación entre las variables observables Ansiedad y Ansiedad social; b) la inclusión de Problemas de atención en el factor Emocional; c) la inclusión de Hiperactividad-impulsividad en el factor Conductual; y d) la inclusión de Problemas de control de la ira en el factor Emocional.

Tabla 4. Estabilidad temporal (fiabilidad test-retest) de las escalas e índices del SENA en la muestra normal.

Escala / Índice	Cuestionario								
	Familia			Escuela			Autoinforme		
	Infantil	Primaria	Secundaria	Infantil	Primaria	Secundaria	Primaria 6-8	Primaria 8-12	Secundaria
DEP	.76	.90	.91	.96	.90	.90	.82	.84	.96
ANS	.89	.90	.89	.9	.89	.87	.85	.83	.93
ASC	.93	.94	.87	.94	.91	.91	---	.88	.92
SOM	.88	.88	.85	.90	.76	.81	---	.87	.90
PST	---	---	---	---	---	---	---	.80	.93
OBS	---	---	---	---	---	---	---	---	.93
ATE	.92	.93	.91	.95	.95	.92	.84	.81	.94
HIP	.91	.91	.94	.91	.94	.92	.90	.84	.89
IRA	.92	.89	.93	.94	.92	.87	.84	.82	.92
AGR	.88	.92	.94	.91	.92	.91	.79	.80	.91
DES	.89	.87	.94	.96	.87	.89	---	.72	.89
ANT	---	---	.87	---	---	.86	---	---	.83
INU	.58	.86	.86	.91	.89	.85	---	---	---
RET	.90	---	---	.97			---	---	---
APR	---	---	---	---	.91	.90	---	---	---

Escala / Índice	Cuestionario								
	Familia			Escuela			Autoinforme		
	Infantil	Primaria	Secundaria	Infantil	Primaria	Secundaria	Primaria 6-8	Primaria 8-12	Secundaria
ALI	---	---	.96	---	---	---	---	---	.95
SUS	---	---	.69	---	---	---	---	---	.95
ESQ	---	---	---	---	---	---	---	---	.92
FAM	---	---	---	---	---	---	.52	.82	.95
ESC	---	---	---	---	---	---	---	.84	.93
COM	---	---	---	---	---	---	.74	.79	.94
REG	.86	.90	.88	.93	.88	.86	---	.75	.92
RIG	.85	.87	.87	.92	.90	.78	---	---	---
AIS	.89	.88	.95	.93	.91	.85	---	---	---
APE	.77	---	---	---	---	---	---	---	---
BUS	---	---	---	---	---	---	---	---	.91
SOC	.87	.92	.90	.92	.90	.91	---	.84	.92
IEM	.86	.87	.82	.92	.89	.88	---	---	---
AUT	---	---	---	---	---	---	.52	.73	.92
CNC	---	---	---	---	---	---	---	---	.90
EST		.91	.91		.90	.91	---	---	---
EMO	.84	.94	.87	.93	.92	.88	.86	.87	.92
CON	.92	.91	.92	.95	.92	.91	.84	.82	.93
EJE	---	---	---	---	---	---	---	.83	.90
CTX	.90	.93	.89	.94	.94	.88	.59	.83	.92
REC	.88	.94	.87	.93	.91	.92	---	.81	.89
GLO	.88	.93	.90	.94	.91	.88	.86	.84	.91

--- : Escala no presente en esa versión del ejemplar del SENA; AGR: Agresión; AIS: Aislamiento; ALI: Problemas de la conducta alimentaria; ANS: Ansiedad; ANT: Conducta antisocial; APE: Dificultades de apego; APR: Problemas de aprendizaje; ASC: Ansiedad social; ATE: Problemas de atención; AUT: Autoestima; BUS: Búsqueda de sensaciones; CNC: Conciencia de los problemas; COM: Problemas con los compañeros; CON: Índice de problemas conductuales; CTX: Índice de problemas contextuales; DEP: Depresión; DES: Conducta desafiante; EJE: Índice de problemas en las funciones ejecutivas; EMO: Índice de problemas emocionales; ESC: Problemas con la escuela; ESQ: Esquizotipia; EST: Disposición al estudio; FAM: Problemas familiares; GLO: Índice global de problemas; HIP: Hiperactividad-impulsividad; IEM: Inteligencia emocional; INU: Comportamiento inusual; IRA: Problemas de control de la ira; OBS: Obsesión-compulsión; PST: Sintomatología postraumática; REC: Índice de recursos personales; REG: Problemas de regulación emocional; RET: Retraso en el desarrollo; RIG: Rigidez; SOC: Integración y competencia social; SOM: Quejas somáticas; SUS: Consumo de sustancias.

Discusión

El SENA se creó como una respuesta actualizada a las demandas que requiere la evaluación de la psicopatología infanto-juvenil, incorporando en su desarrollo importantes características para el contexto clínico y educativo: 1) diferentes versiones ajustadas a criterios evolutivos y dirigidas tanto al propio niño como a sus padres y profesores (permite obtener información de diferentes fuentes y contextos); 2) evaluación de múltiples dimensiones que cubran las principales alteraciones psicopatológicas y los síntomas más relevantes, áreas de vulnerabilidad, recursos personales y niveles de adaptación al contexto familiar y escolar; y 3) una aproximación dimensional a la evaluación psicopatológica, que permite identificar no solo la presencia de las alteraciones sino el grado de intensidad con que se manifiestan al tiempo que se apoya en criterios diagnósticos actualizados (APA, 2013).

Durante su construcción, se ha seguido un estricto proceso de diseño y desarrollo, ajustándose a las directrices internacionales sobre el uso y desarrollo de los test (Abad et al., 2011; AERA, APA, & NCME, 2014; Downing & Haladyna, 2010; ITC, 2001, 2014; Navas, 2001; Schmeiser & Welch, 2006), incluyendo la revisión de la literatura científica disponible, la generación de tablas de especificaciones de los ítems, su evaluación por parte de expertos de distintas áreas y especialidades, la realización de estudios piloto o la consideración de múltiples indicadores para la toma de decisiones sobre cada uno de los ítems y escalas que constituyen la versión definitiva del SENA.

Las puntuaciones obtenidas presentan coeficientes de consistencia interna satisfactorios o muy satisfactorios en la inmensa mayoría de las escalas e índices (> .70) de acuerdo a los estándares habitualmente establecidos (Hernández, Ponsoda, Muñiz, Prieto, & Elosua, 2016). Los coeficientes de fiabilidad test-retest apuntan en la misma dirección e indican que la estabilidad temporal de las mismas es tam-

Tabla 5. Índices de ajuste de los modelos contrastados en el AFC de cada cuestionario del SENA.

Cuestionario	Índices de ajuste			
	CFI	TLI	RMSEA	SRMR
Autoinforme primaria 6-8 años				
Modelo de 1 factor	.782	.637	.237	.109
Modelo de 2 factores	.925	.859	.148	.065
Modelo de 3 factores	.969	.923	.109	.044
Modelo de 3 factores modificado	.986	.965	.074	.034
Autoinforme primaria 8-12 años				
Modelo de 1 factor	.803	.747	.184	.095
Modelo de 2 factores	.926	.902	.115	.061
Modelo de 3 factores	.959	.942	.088	.050
Modelo de 3 factores modificado	.974	.962	.071	.042
Autoinforme secundaria				
Modelo de 1 factor	.670	.596	.197	.136
Modelo de 2 factores	.870	.838	.125	.075
Modelo de 3 factores	.907	.880	.107	.065
Modelo de 3 factores modificado	.926	.902	.097	.055
Familia infantil				
Modelo de 1 factor	.746	.661	.176	.103
Modelo de 2 factores	.807	.733	.157	.091
Modelo de 3 factores	.888	.832	.124	.082
Modelo de 3 factores modificado	.921	.877	.106	.059
Familia primaria				
Modelo de 1 factor	.692	.589	.208	.122
Modelo de 2 factores	.845	.786	.150	.076
Modelo de 3 factores	.917	.876	.114	.064
Modelo de 3 factores modificado	.951	.923	.090	.042
Familia secundaria				
Modelo de 1 factor	.734	.659	.200	.116
Modelo de 2 factores	.853	.805	.151	.080
Modelo de 3 factores	.899	.858	.129	.069
Modelo de 3 factores modificado	.931	.900	.108	.050
Escuela infantil				
Modelo de 1 factor	.641	.522	.235	.152
Modelo de 2 factores	.788	.707	.184	.110
Modelo de 3 factores	.886	.828	.141	.114
Modelo de 3 factores modificado	.905	.852	.131	.099
Escuela primaria				
Modelo de 1 factor	.629	.505	.255	.156
Modelo de 2 factores	.785	.702	.198	.108
Modelo de 3 factores	.866	.799	.163	.100
Modelo de 3 factores modificado	.906	.854	.139	.090
Escuela secundaria				
Modelo de 1 factor	.678	.586	.170	.105
Modelo de 2 factores	.689	.588	.170	.105
Modelo de 3 factores	.841	.777	.125	.078
Modelo de 3 factores modificado	.839	.767	.128	.079

CFI: Índice de ajuste comparativo; TLI: Índice de Tucker-Lewis; RMSEA: Raíz media cuadrática del error de aproximación; SRMR: Residuo cuadrático medio estandarizado.

bién muy satisfactoria (> .70) e incluso elevada dada la edad a la que se dirige la prueba (del Barrio, 2014). Estas evidencias representan un apoyo a la precisión de las puntuaciones obtenidas a partir de las escalas e índices del SENA.

Los resultados de los análisis factoriales confirmatorios evidencian que la estructura de los datos obtenidos con la aplicación del instrumento se ajusta en gran medida a la estructura teórica prevista y suponen un apoyo a la interpretación de los índices obtenidos a partir de las escalas principales. Los índices de ajuste obtenidos son concordantes con los obtenidos en la literatura científica en otros instrumentos de esta complejidad que recogen una gran heterogeneidad de constructos, con diversidad de informantes y donde el foco de atención se encuentra en las escalas de la prueba y los índices factoriales resultan de interés secundario (p. ej., BASC y BASC-2; Reynolds & Kamphaus, 2004a, 2004b). Es relevante destacar la agrupación de las diferentes escalas en dos grandes dimensiones de amplio espectro, la dimensión interiorizada y la exteriorizada que el SENA extrae como índices globales de problemas y que resulta coincidente con las estructuras previas de otros instrumentos multidimensionales dirigidos a la evaluación de la psicopatología infantojuvenil (Achenbach & Rescorla, 2000, 2001; Reynolds & Kamphaus, 2004a, 2004b). Estas dos dimensiones globales se complementan con un tercer factor comprehensivo, no siempre presente en los instrumentos previos, que es la dimensión de función ejecutiva. Todo ello permite una estructura acorde al mapa de manifestaciones sintomáticas más prevalentes en la población infantojuvenil: ansiedad y depresión, problemas de conducta y problemas atencionales y de hiperactividad (Ezpeleta et al., 2007; Polanczyk et al., 2015). Esta estructura interna del SENA le confiere validez de constructo y utilidad para su uso en niños y adolescentes.

Dado el carácter multidimensional del SENA, el elevado número de constructos que evalúa y las diferencias evolutivas que, como recomiendan especialistas en evaluación infantil (Forns, 1993), fueron diseñadas para cada informador y nivel de edad, por limitaciones de extensión, en este artículo solo se han podido presentar algunas evidencias relacionadas con la validez del instrumento, en particular con su estructura interna. Otras informaciones, como las intercorrelaciones de las escalas y las evidencias externas relacionadas con la convergencia con otras pruebas similares o la capacidad diagnóstica de las escalas en diferentes grupos clínicos se pueden consultar en Fernández-Pinto et al. (2015a, 2015b).

Una de las características más positivas del SENA es que los datos obtenidos mediante su aplicación permiten tener en cuenta no solo el diagnóstico de los problemas, sino también la prevención de los mismos y los recursos que existen para llevar a cabo la prevención primaria y secundaria de los problemas, cosa que es hoy característica imprescindible de una buena evaluación (Ordoñez, Maganto, & González, 2014).

Por otra parte el nivel de problematicidad alcanzado en cada una de las distintas escalas del SENA es homogéneo y acorde con los estudios epidemiológicos que se vienen encontrando en distintos países (Rescorla, Achenbach, Ivanova, Dumenci, Almqvist et al., 2007) lo que es, en definitiva una validación externa.

En conclusión, las evidencias disponibles sobre el SENA resultan favorables y apoyan su uso para los fines para los que fue creado. Esto permite contar con un nuevo instrumento de evaluación con adecuadas garantías psicométricas y que proporciona información de manera rápida, eficiente, fácil y no invasiva sobre los problemas emocionales y conductuales en niños y adolescentes.

Artículo recibido: 09/05/2016
Aceptado: 14/06/2016

Conflicto de intereses

Los autores declaran no tener conflicto de intereses.

Referencias

Abad, J. F., Olea, J., Ponsoda, V., & García, C. (2001). *Medición en ciencias sociales y de la salud*. Madrid: Editorial Síntesis.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213–232. doi:10.1037/0033-2909.101.2.213

Alison L. Neil, A. L., & Christensen, H. (2009). Efficacy and effectiveness of school-based prevention and early intervention programs for anxiety. *Clinical Psychology Review*, *29*(3), 208–215. doi:10.1016/j.cpr.2009.01.002

American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* Washington, DC: AERA Publications.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (5.ª Ed.)*. Washington, DC: APA.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7.ª ed.). Upper Saddle River, NJ: Prentice Hall. Edición Española: Tests psicológicos. Madrid: Prentice Hall.

Barkley, R.A. (2006). Symptoms, diagnosis, prevalence and gender differences. En R. A. Barkley (Ed.), *Attention-Deficit Hyperactivity Disorder. A Handbook for diagnosis and treatment*. (3.ª ed.) (pp. 77-121). Londres: The Gilford Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M. (1995). *EQS, Structural Equations Program Manual*. Encino, CA: Multivariate Software.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. En K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage Publications.

Calear, A. L., & Christensen, H. (2010). Systematic review of school-based prevention and early intervention programs for depression. *Journal of Adolescence*, *33*(3), 429–438. doi:10.1016/j.adolescence.2009.07.004

Carrasco, M. A. (2013). La evaluación clínica: introducción a los aspectos conceptuales y técnicos básicos. En M. A. Carrasco, I. Ramírez, & M. V. del Barrio (Eds.), *Evaluación Clínica. Diagnóstico, formulación y contrastación de los trastornos psicológicos* (pp. 3-40). Madrid: Sanz y Torres.

Cohen, B. H. (1996). *Explaining Psychological Statistics*. Pacific Grove, CA: Brooks/Cole.

Conners, C. K. (2008). *CONNERS. Comprehensive Behavior Rating Scales*. Toronto, ON: Multi-Health Systems.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research and Evaluation, 10*(7), 1-9. Recuperado de http://pareonline.net/getvn.asp?v=10&n=7

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16(3)*, 297-334.

De Los Reyes, A., & Kazdin, A. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483–509. doi:10.1037/0033-2909.131.4.483

De Pauw, S. S. W., & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: a review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry Human Development, 41*(3), 313-329. doi:10.1007/s10578-009-0171-8

del Barrio, V. (2002a). *Evaluación psicológica en la infancia y la adolescencia*. Madrid: UNED.

del Barrio, V. (2002b). Método de la evaluación psicológica y su aplicación al niño y adolescente. En M. V. del Barrio (Coord.), *Evaluación psicológica en la infancia y la adolescencia* (pp. 57-88). Madrid: UNED.

del Barrio,V. (2014). *Evaluación Psicológica*. Madrid: Centro de Estudios Financieros.

Downing, S. M., & Haladyna, T. M. (2010). *Handbook of Test Development*. Nueva York: Routledge.

Ezpeleta, L. J. (2014). Prevalence of DSM-IV disorders, comorbidity and impairment in 3-year-old Spanish preschoolers. *Social Psychiatry and Psychiatric Epidemiology, 49*(1), 145-155. doi:10.1007/s00127-013-0683-1

Ezpeleta, L., Guillamón, N., Granero, R., de la Osa, N., Domènech, J. M., & Moya, I. (2007). Prevalence of mental disorders in children and adolescents from a Spanish slum. *Social Science and Medicine, 64*(4), 842-849. doi:10.1016/j.socscimed.2006.10.031

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299. doi:10.1037/1082-989X.4.3.272

Fernández-Ballesteros, R., De Bruyn, E., Godoy, A., Hornke, L., Ter Laak, J., Vizcarro, C., ... & Zaccagnini, J. L. (2003). Guías para el proceso de evaluación (GAP): Una propuesta a discusión. *Papeles del Psicólogo, 84*, 58-70.

Fernández-Pinto, I., Santamaría, P., Sánchez-Sánchez, F., Carrasco, M. A., & del Barrio, V. (2015a). *SENA. Sistema de Evaluación de Niños y Adolescentes. Manual técnico*. Madrid: TEA Ediciones.

Fernández-Pinto, I., Santamaría, P., Sánchez-Sánchez, F., Carrasco, M. A., & del Barrio, V. (2015b). *SENA. Sistema de Evaluación de Niños y Adolescentes. Manual de aplicación, corrección e interpretación*. Madrid: TEA Ediciones.

Frick, P. J., & Nigg, J. T. (2012). Current Issues in the Diagnosis of Attention Deficit Hyperactivity Disorder, Oppositional Defiant Disorder, and Conduct Disorder. *Annual Review of Clinical Psychology, 8*, 77–107. doi:10.1146/annurev-clinpsy-032511-143150

Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G., & Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo, 37,* 161-168.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. doi:10.1080/10705519909540118

Hunsley, J., & Mash, E. J. (2007). Evidence-Based Assessment. *Annual Review Clinical Psychology, 3*, 29–51.

IBM Corp. (2010). *IBM SPSS Statistics for Windows, Version 19.0*. Armonk, NY: IBM Corp.

Instituto Nacional de Estadística (2014). *Padrón. Población por municipio.* Recuperado de http://www.ine.es

International Test Commision (2014). ITC Guidelines on Quality Control in Scoring, Test Analysis, and Reporting of Test Scores. *International Journal of Testing*, *14*(3), 195-217. doi:10.1080/15305058.2014.918040

International Test Commission (2001). International Guidelines for Test Use. *International Journal of Testing*, *1*(2), 93-114. doi:10.1207/S15327574IJT0102_1

Jensen, P. S., Hinshaw, S. P., Kraemer, H. C., Lenora, N., Newcorn, J. H., Abikoff, H. B., ... Vitiello, B. (2001). ADHD comorbidity findings from the MTA study: comparing comorbid subgroups. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(2), 147-58. doi:10.1097/00004583-200102000-00009

Kamphaus, R. W., & Frick, P. J. (1996). *Clinical assessment of child and adolescent personality and behavior*. Needham Heights, MA: Allyn & Bacon.

Kendall, P. C. (2012). *Child and Adolescents Therapy.* Nueva York: the Guilford Press.

Lachar, D., & Gruber, C. P. (1995). *Personality Inventory for Youth (PIY).* Los Ángeles, CA: WPS.

Lachar, D., & Gruber, C. P. (2001). *Personality Inventory for Children – 2.ª Edición (PIC-2)*. Los Ángeles, CA: WPS.

Lachar, D., Wingenfeld, S. A., Kline, R. B., & Gruber, C. P. (2000). *Student Behavior Survey (SBS).* Los Ángeles, CA: WPS.

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research, 34,* 347-365. doi:10.1207/S15327906MBR3403_3

Lorenzo-Seva, U., & Ferrando, P. J. (2012). *Factor 9.0*. Tarragona: Universidad Rovira i Virgili.

Martínez-Arias, R. (1995). *Psicometría: Teoría de los tests psicológicos y educativos.* Madrid: Síntesis Psicología.

Martínez-Arias, R., Hernández-Lloreda, M. J., & Hernández-Lloreda, M. V. (2006). *Psicometría.* Madrid: Alianza Editorial.

Muthén, B., & Muthén, L. K. (2006). *Mplus 4.1*. Los Ángeles: Muthén & Muthén.

Navarro, E., Moral, J. C., Galán, A., & Beitia, M. D. (2012). Desarrollo infantil y adolescente: Trastornos mentales más frecuentes en función de la edad y el género. *Psicothema, 24*, 377-383.

Navas, M. J. (2001). *Métodos, diseños y técnicas de investigación psicológica*. Madrid: UNED.

Ordoñez, A., Maganto, C., & González, R. (2014). Quejas somáticas, conciencia emocional e inadaptación en población escolar. *Anales de Pediatría, 85*, 308-331. doi:10.1016/j.anpedi.2014.03.020

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry, 56*(3), 345-365. doi:10.1111/jcpp.12381

Rescorla, L., Achenbach, Th. M., Ivanova, M. Y., Dumenci, L., Almqvist, F., et al. (2007). Epidemiological comparisons of problems and positive qualities reported by adolescents in 24 countries. *Journal of Consulting and Clinical Psychology, 75,* 351-358. doi:10.1037/0022-006X.75.2.351

Reynolds, C. R. (2010). Measurement and assessment: An editorial view. *Psychological Assessment, 22*, 1-4. doi:10.1037/a0018811

Reynolds, C. R., & Kamphaus, R. W. (2004a). *BASC-2. Behavior Assessment System for Children – 2.ª Edición*. Bloomington, MN: Pearson.

Reynolds, C. R., & Kamphaus, R. W. (2004b). *BASC. Sistema de Evaluación de la Conducta de Niños y Adolescentes* (J. González, S. Fernández, E. Pérez, & P. Santamaría, adaptadores). Madrid: TEA Ediciones.

Schmeiser, C. B., & Welch, C. J. (2006). Test development. En R. L. Brennan (Ed.), *Educational Measurement*, (vol. 4, pp. 307-353). Wesport, CT: National Council on Measurement in Education and American Council on Education.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173- 180. doi:10.1207/s15327906mbr2502_4

Tucker, L. R., & Lewis, C. (1973). The reliability coeficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10. doi:10.1007/BF02291170

Unitat d'Epidemiologia i de Diagnòstic en Psicopatologia del Desenvolupament (2013). *Baremos para CBCL 6-182001. Población española*. Barcelona: Universidad Autónoma de Barcelona.