

# Evaluación de

## Sistema de evaluación de la conducta de niños y adolescentes - 3

∞ **BASC-3** ∞

# RESUMEN DE LA VALORACIÓN DEL TEST

## Descripción general

Característica	Descripción
Nombre del test	Sistema de evaluación de la conducta de niños y adolescentes - 3 (BASC 3)
Autor	Cecil R. Reynolds; Randy W. Kamphaus
Autor de la adaptación española	Departamento de I+D de Pearson Clinical & Talent Assessment: Ana Hernández, Èrica Paradell y Frédérique Vallar
Variable(s)	Problemas conductuales y emocionales de los niños y adolescentes.
Áreas de aplicación	Psicología clínica Psicología educativa Psicología forense
Soporte	Para su aplicación, papel y lápiz o plataforma online. Para su corrección, plataforma online.

## Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Buena-Excelente	4.8
Fundamentación teórica	Buena	4
Adaptación	Adecuada - Buena	3.5
Análisis de ítems	No aporta información	-
Validez: contenido	Buena	4
Validez: relación con otras variables	Adecuada	3
Validez: estructura interna	Adecuado con carencias	2
Validez: análisis del DIF	-	-
Fiabilidad: equivalencia	-	-
Fiabilidad: consistencia interna	Buena-Excelente	4.5
Fiabilidad: estabilidad	Adecuada-Buena	3.5
Fiabilidad: TRI	-	-
Fiabilidad inter-jueces	-	-
Baremos e interpretación de puntuaciones	Buena	4

## Comentarios generales

La BASC3 es un sistema multimétodo y multifuente para evaluar dificultades conductuales y emocionales de niños y adolescentes de 3 a 18 años. Cuenta con la ventaja de presentar tres formas: cuestionario para profesores/tutores, cuestionario para padres y cuestionario autoinforme, lo que permite contemplar su comportamiento en diferentes contextos (i.e., familiar, escolar) y desde diferentes perspectivas (i.e., la del propio menor, la de los padres y la de los profesores). Destaca también la variedad de escalas (clínicas, adaptativas, de contenido, compuestas) e índices (clínicos y de funcionamiento ejecutivo), que proporcionan una valoración bastante exhaustiva. Esta información global puede ser teóricamente muy útil para el profesional educativo, clínico, o forense, permitiendo identificar cuáles son las vulnerabilidades y potencialidades del menor y en qué contextos, lo que facilita la toma de decisiones sobre la posible necesidad de realizar una evaluación más profunda sobre alguno de los aspectos, y cuáles son las intervenciones más adecuadas a realizar, etc.

La BASC es un instrumento de evaluación con larga tradición y prestigio, siendo la BASC-3, la última versión. En la adaptación de la prueba ha participado un amplio grupo de expertos, entre ellos una colaboradora en la adaptación española de la primera versión del BASC; se ha contado con la colaboración estrecha del autor original de la prueba, Randy Kamphaus, y con el apoyo del equipo internacional de Pearson Clinical & Talent Assessment. Se ha prestado especial atención al proceso de traducción y adaptación de la prueba, atendiendo a los siguientes criterios: (a) nivel de lectura equivalente o inferior al de 2º de primaria; (b) relevancia del contenido del ítem para la escala de medida; (c) adecuación cultural; y (d) adecuación al rango de edad del cuestionario.

Las tres formas de la BASC3 pueden aplicarse en papel o de forma online, aunque todas se corrigen mediante una plataforma digital web, lo que facilita a los profesionales obtener las puntuaciones y percentiles, sin necesidad de hacer ningún cálculo. Una fortaleza de la prueba es la obtención de una muestra de tipificación compuesta de 1114 aplicaciones (342, en el caso del cuestionario T; 480 en el caso del cuestionario P; y 322, en el caso del cuestionario S), cuidando que fuera representativa de la población a la que va dirigida la prueba. La muestra era estratificada, manteniendo representación por sexo, edad y nivel educativo de los padres. También se ha puesto especial cuidado en la obtención de participantes pertenecientes a todas las comunidades autónomas de España.

Se proporcionan baremos separando siempre por edad, con posibilidad de obtener baremos generales o específicos de cada sexo. En cualquiera de los casos se ofrecen las puntuaciones T y el percentil del niño o adolescente evaluado. Además, se ofrece una clasificación según las puntuaciones obtenidas en las escalas adaptativas y escalas clínicas, valorando el nivel de conductas problemáticas y adaptativas. La prueba cuenta también con diversos índices que permiten analizar la calidad de las respuestas de cada sujeto, lo que constituye también una fortaleza.

La fiabilidad (tanto la consistencia interna como la estabilidad) de las puntuaciones es generalmente buena o excelente en las tres formas de la prueba y para todos los grupos de edad. Más allá del cuidado proceso de desarrollo de la prueba, para la muestra española se aportan principalmente tres tipos de evidencias de validez, que apoyan la validez de algunas de las escalas e índices:

1) Matrices multirasgo-multimétodo para evaluar la divergencia/convergencia de puntuaciones en los distintos cuestionarios, encontrando resultados consistentes con los hallados en estudios previos,

2) Análisis factorial para justificar la composición de las escalas compuestas. La evidencia basada en la estructura interna muestra que las correlaciones entre las escalas clínicas y las correlaciones entre las escalas adaptativas son positivas, mientras que las correlaciones entre las escalas clínicas y las escalas adaptativas son negativas. El análisis factorial de las escalas que componen las escalas compuestas (mediante el método de ejes principales) respalda el agrupamiento propuesto en la versión estadounidense para varias de las escalas (p.ej., las del cuestionario T1).

3) Diferencias de puntuaciones entre el grupo general y grupos especiales de niños y adolescentes con determinados diagnósticos clínicos (TDAH, TEA). Con relación a estas últimas evidencias, se observa el patrón esperado. Para el grupo TDAH se esperan puntuaciones más altas que las de la población general en las escalas Problemas de atención e Hiperactividad. Para el grupo con TEA se esperan puntuaciones altas en varias escalas compuestas y clínicas (p.ej., Atipicidad), así como puntuaciones considerablemente por debajo del rango promedio en las

escalas adaptativas.

Algunos aspectos mejorables son:

1) Algunas evidencias (p.ej., correlaciones con otras variables, o fiabilidad test-retest para el cuestionario T) sólo se han obtenido en muestra estadounidense. Debe tenerse en cuenta que las evidencias mostradas no pueden aplicarse directamente a la adaptación española, sobre todo considerando que en esta adaptación se han modificado diversos ítems, incluso en su contenido, por lo que convendría obtener ese tipo de evidencia en la muestra española.

2) Aunque la muestra global es amplia y representativa, los tamaños muestrales disponibles para cada versión del cuestionario son más reducidos. Por ejemplo, el promedio de los tamaños muestrales para las distintas versiones (T1, T2, T3, P1, P2, P3, S2 y S3) es 143. Por tanto, la muestra final para cada análisis es más reducida. Este problema afecta a todas las evidencias recogidas de fiabilidad y validez, por lo que convendría ampliar la muestra para estos análisis.

3) Con relación a las evidencias de fiabilidad. Sería recomendable incluir algún estudio de fiabilidad interjueces, para los cuestionarios T y P, para valorar el efecto que tiene el informante en el error de medida. Por último, en el caso de que se ampliara la muestra, el uso de la Teoría de la Respuesta al Ítem permitiría obtener la información para cada nivel de rasgo y, con la presencia de ítems de anclaje entre cuestionarios, estimar las puntuaciones a los distintos cuestionarios en la misma métrica. Un análisis conjunto de las distintas escalas permitiría valorar en qué grado las distintas versiones del cuestionario están midiendo los mismos constructos.

4) Con relación a las evidencias de validez de estructura interna. Aunque las escalas clínicas, adaptativas y de contenido se han construido siguiendo criterios clínicos y de sensibilidad diagnóstica, con una fuerte validez de contenido, sería conveniente proporcionar evidencia, para la muestra española, sobre la composición/estructura de las escalas adaptativas y clínicas (sólo se realizan análisis factoriales para validar la composición de las escalas compuestas). Por otro lado, muchas de las escalas adaptativas y clínicas incluidas en el análisis para validar la composición de las escalas compuestas saturan en varios factores, lo que da lugar a una estructura difícil de interpretar, considerando lo que evalúa este cuestionario.

5) Con relación a las evidencias de validez de relación con otras variables. Sería recomendable que se informara de si las diferencias entre los grupos son significativas, así como de los tamaños del efecto.

6) Puntos de corte. De acuerdo con el manual, las puntuaciones no deben utilizarse para hacer diagnósticos específicos, ya que para ello es necesario utilizar otras técnicas de evaluación (i.e., entrevista estructurada, observación, etc.). No obstante, también se indica que los componentes del BASC-3 pueden servir de ayuda para el diagnóstico clínico. En una escala de este tipo (i.e., en las que se clasifica a los evaluados en niveles de "en riesgo" o "clínicamente significativo"), sería conveniente incluir más evidencia que justifique los puntos de corte utilizados (que, en este caso, se basan en el criterio estadístico -alejamiento de la media).

7) Baremos. Al diferenciar simultáneamente por edad y sexo, los tamaños son demasiado pequeños. Por ello en el manual se recomienda utilizar los baremos combinados por sexo, ya que en la mayoría de los casos permiten llevar a cabo un diagnóstico más preciso. Sería recomendable ampliar la muestra de baremación en el futuro.

8) Manual. Aunque en el manual se explican algunos aspectos en profundidad, hay otros que requerirían mayor detalle. Por ejemplo, en el manual se explica que se realizaron análisis factoriales para justificar la composición de los índices de funcionamiento ejecutivo, pero sólo se proporcionan los índices de ajuste (con valores inadecuados en algunos casos) y no se muestran las saturaciones ni se proporciona ninguna otra información al respecto. Por otra parte, en los análisis factoriales de las escalas compuestas no se muestran los índices de ajuste, que podrían haberse obtenido de utilizar otro método de estimación distinto a ejes principales, ni tampoco las correlaciones entre factores. Y en el caso de las correlaciones con otros test, en muestra estadounidense, no se proporciona ninguna información sobre las características de estos cuestionarios. Tampoco se proporciona ninguna información del análisis de ítems de las escalas.

En resumen, debe reconocerse que la BASC es una prueba con una trayectoria sólida en cuanto a su uso y que, conceptualmente, es un cuestionario útil y necesario. Sería recomendable recoger más evidencia sobre la validez de la BASC-3 en muestra española. Se debería aportar más información sobre la estructura factorial de las escalas (p.ej., clínicas y adaptativas), el funcionamiento diferencial de los ítems, así como sobre las correlaciones con otros tests y criterios.

# ANÁLISIS DETALLADO DE LA PRUEBA

## 1. DESCRIPCIÓN GENERAL DEL TEST

### 1.1. Nombre del test:

Sistema de evaluación de la conducta de niños y adolescentes - 3 (BASC 3)

### 1.2. Nombre del test en su versión original:

Behavior Assessment System for Children, Third Edition

### 1.3. Autor/es del test original:

Cecil R. Reynolds; Randy W. Kamphaus

### 1.4. Autor/es de la adaptación española:

Departamento de I+D de Pearson Clinical & Talent Assessment: Ana Hernández, Èrica Paradell y Frédérique Vallar

### 1.5. Editor del test en su versión original:

NCS Pearson Inc

### 1.6. Editor de la adaptación española:

Pearson Educación

### 1.7. Fecha de publicación del test original:

2015

### 1.8. Fecha de publicación del test en su adaptación española:

2020

### 1.9. Fecha de la última revisión del test:

--

### 1.10. Área general de la/s variable/s que pretende medir el test:

Personalidad, Escalas clínicas, Problemas de conducta y emocionales.

## 1.11. Breve descripción de la/s variable/s que pretende medir el test:

El instrumento "BASC-3. Sistema de Evaluación de la conducta de niños y adolescentes" es un sistema multimétodo y multidimensional para evaluar problemas conductuales y emocionales, considerando sus manifestaciones en diferentes actividades y entornos sociales, en menores de 3 a 18 años. Consta de cuestionarios para profesores/tutores (T), cuestionarios para padres (P) y cuestionarios autoinforme (S). Por lo tanto, permite recoger información de diferentes fuentes, para alcanzar una mejor comprensión tanto de los problemas como de los puntos fuertes de los menores.

Los cuestionarios dan lugar a cuatro tipos de escalas (clínicas, adaptativas, de contenido, compuestas) y a dos tipos de índices (clínicos y de funcionamiento ejecutivo):

- Escalas clínicas (p.ej., Agresividad, Ansiedad). Miden conductas desadaptativas. Niveles altos indican niveles problemáticos de funcionamiento.

- Escalas adaptativas (p.ej., Adaptabilidad, Habilidades sociales). Miden conductas adaptativas o puntos fuertes. Puntuaciones bajas indican posibles áreas problemáticas.

- Escalas de contenido (p.ej., Resiliencia, Acoso escolar) comparten algunos ítems con las subescalas clínicas y adaptativas, y se entienden como más generales.

- Escalas compuestas (p.ej., Problemas de exteriorización, Habilidades adaptativas). Se obtienen también de las subescalas clínicas y adaptativas, y se entienden como factores de orden superior, habiéndose derivado en parte de los resultados del análisis factorial.

- Índices clínicos (p.ej., de autismo) y de funcionamiento ejecutivo (p.ej., de control atencional). Se derivan de ítems de otras escalas y están vinculados con el diagnóstico de síntomas conductuales o problemas de funcionamiento emocional. Los ítems de estas escalas se han seleccionado empíricamente por su capacidad de diferenciar niños y adolescentes que presentan un diagnóstico o clasificación conductual o emocional de aquellos que no lo presentan.

Como se ha mencionado, existen tres tipos de cuestionarios, que a continuación se describen brevemente.

- Cuestionarios T y P.

Los cuestionarios T y P varían en función de la edad del menor: hay una versión para los niños de 3-5 años (T1 y P1), otra para 6-11 años (T2 y P2) y otra para 12-18 años (T3 y P3). Estos cuestionarios pueden cumplimentarse online o en papel en unos 10-20 minutos. Las formas T y P incluyen:

a) Escalas clínicas (entre 8 y 10 escalas según la forma):

- Agresividad
- Ansiedad
- Atipicidad
- Depresión
- Hiperactividad
- Problemas de aprendizaje (en T2 y T3)
- Problemas de atención
- Problemas de conducta (en P2, P3, T2 y T3)
- Retraimiento
- Somatización.

b) Escalas adaptativas (entre 4 y 5 escalas, según la forma):

- Actividades cotidianas (en P1, P2 y P3)
- Adaptabilidad
- Comunicación funcional
- Habilidades académicas (en T2 y T3)
- Habilidades sociales
- Liderazgo (en P2, P3, T2 y T3)

c) Escalas de contenido (7 escalas):

- Acoso escolar
- Autocontrol emocional
- Control de la ira
- Emocionalidad negativa
- Funcionamiento ejecutivo
- Resiliencia
- Trastornos del desarrollo social

d) Escalas compuestas (entre 4 o 5 escalas, según la forma; entre paréntesis se muestran las escalas clínicas y adaptativas a partir de las que se obtienen, si bien la composición exacta depende de la versión del cuestionario):

- Problemas de exteriorización (Hiperactividad, Agresividad y Problemas de conducta)
- Problemas de interiorización (Ansiedad, Depresión y Somatización)
- Habilidades adaptativas (Adaptabilidad, Actividades cotidianas, Comunicación funcional, Habilidades sociales, Liderazgo y Habilidades académicas)
- Índice de síntomas conductuales (Hiperactividad, Agresividad, Depresión, Problemas de atención, Atipicidad y Retraimiento).
- Problemas escolares (Problemas de aprendizaje y Problemas de atención), en las formas T2 y T3.

e) Índices. Los índices proporcionados dependen de la edad del evaluado. Se incluyen índices clínicos (de deterioro funcional, de probabilidad clínica, de autismo, de comportamiento perturbador y de probabilidad de TDAH) y de funcionamiento ejecutivo (de control atencional, de control conductual, de control emocional, de funcionamiento ejecutivo general y de resolución de problemas).

- Cuestionarios S.

El cuestionario S puede cumplimentarse online con supervisión o en papel en unos 20-30 minutos. El cuestionario de autoinforme permite que los niños y adolescentes registren su mundo interior (sus pensamientos, sentimientos, actitudes y reacciones internas respecto a personas y hechas). En el caso de la forma S, tan sólo hay una versión para 8-11 años (niños; S2) y otra para 12-18 años (adolescentes; S3), dado que los niños de 3-7 años son demasiado pequeños como para rellenar un cuestionario autoinforme. La forma S incluyen:

a) Escalas clínicas (entre 10 y 12 escalas, según la forma):

- Actitud respecto a la escuela
- Actitud respecto a los profesores
- Ansiedad
- Atipicidad
- Búsqueda de sensaciones (sólo en S3)
- Depresión
- Estrés social
- Hiperactividad
- Locus de control
- Problemas de atención
- Sensación de ineptitud
- Somatización (sólo en S3).

b) Escalas adaptativas (4 escalas):

- Autoestima
- Confianza en sí mismo
- Relación con los padres



- Relaciones interpersonales.

c) Escalas de contenido (4 escalas; sólo en S3):

- Ansiedad ante los exámenes

- Control de la ira

- Fuerza del ego

- Manía

d) Escalas compuestas (5 escalas; entre paréntesis se muestran las escalas clínicas y adaptativas a partir de las que se obtienen, si bien la composición exacta depende de la versión del cuestionario):

- Problemas escolares (Actitud respecto a la escuela, Actitud respecto a los profesores y Búsqueda de sensaciones)

- Problemas de interiorización (Atipicidad, Locus de control, Estrés social, Ansiedad, Depresión, Sensación de ineptitud y Somatización)

- Falta de atención/hiperactividad (Problemas de atención e Hiperactividad)

- Índice de síntomas emocionales (Estrés social, Ansiedad, Depresión, Sensación de ineptitud, Autoestima y Confianza en sí mismo)

- Ajuste personal (Relación con los padres, Relaciones interpersonales, Autoestima y Confianza en sí mismo).

e) Índices. Se proporciona un único índice clínico, denominado índice de deterioro funcional.

- Escalas de validez.

Adicionalmente, estos cuestionarios P, T y S permiten calcular una serie de índices que permiten detectar patrones de respuesta anómalos o inconsistentes.

En resumen, estos cuestionarios contienen ítems relacionados con diferentes aspectos, como personalidad, problemas emocionales, problemas conductuales, problemas relacionados con el entorno académico (ansiedad ante los exámenes, actitud respecto a los profesores, etc.) o habilidades, entre otros. El motivo es que pretende ofrecer una imagen global tanto de los problemas como de los puntos fuertes del menor.

### 1.12. Áreas de aplicación:

Psicología clínica, Psicología educativa, Psicología forense

### 1.13. Formato de los ítems:

En los cuestionarios T y P: Respuesta graduada (formato de frecuencia de cuatro categorías ordenadas: Nunca, A veces, Frecuentemente, Casi siempre).

En los cuestionarios S: Verdadero/Falso y respuesta graduada (formato de frecuencia de cuatro categorías ordenadas: Nunca, A veces, Frecuentemente, Casi siempre).

### 1.14. Número de ítems:

- Por tipo de cuestionario:

- Cuestionario T (profesores/tutores):

- T1, preescolar: 105 ítems

- T2, niños: 156 ítems

- T3, adolescentes: 165 ítems

- Cuestionario P (para padres):

- P1, preescolar: 139 ítems

- P2, niños: 175 ítems
- P3, adolescentes: 173 ítems
- Cuestionarios S (Autoinforme):
- S2, niños: 137 ítems
- S3, adolescentes: 189 ítems

- Número de ítems para las escalas compuestas, clínicas y adaptativas (en cursiva aparecen las escalas clínicas/adaptativas que forman parte de varias escalas compuestas):

<b>Escalas compuestas (y clínicas/adaptativas que las componen)</b>	T1	T2	T3	P1	P2	P3	S2	S3
Compuesta: Problemas de exteriorización (en P y T)	19	30	32	19	30	32		
*Clínica: Agresividad	10	10	11	8	9	10		
*Clínica: Hiperactividad	9	11	11	11	11	8		
*Clínica: Problemas de conducta		9	10		10	14		
Compuesta: Problemas de interiorización (en P, T y S)	26	27	28	35	39	36	56	73
*Clínica: Ansiedad	9	9	9	11	14	13	11	13
*Clínica: Depresión	10	11	12	11	13	13	10	12
*Clínica: Somatización.	7	7	7	13	12	10		7
*Clínica: Atipicidad							10	10
*Clínica: Estrés social							9	11
*Clínica: Locus de control							8	8
*Clínica: Sensación de ineptitud							8	12
Compuesta: Habilidades adaptativas (en P y T)	22	44	45	36	46	46		
*Adaptativa: Actividades cotidianas				7	9	8		
*Adaptativa: Adaptabilidad	7	9	8	9	8	8		
*Adaptativa: Comunicación funcional	9	10	9	11	12	12		
*Adaptativa: Habilidades académicas		8	11					
*Adaptativa: Habilidades sociales	6	10	10	9	10	10		
*Adaptativa: Liderazgo		7	7		7	8		
Compuesta: Índice de síntomas conductuales (en P y T)	51	56	62	59	63	59		
*Clínica: Agresividad	10	10	11	8	9	10		
*Clínica: Hiperactividad	9	11	11	11	11	8		
*Clínica: Depresión	10	11	12	11	13	13		
*Clínica: Atipicidad	8	8	10	10	14	11		
*Clínica: Problemas de atención	7	8	10	7	7	9		
*Clínica: Retraimiento	7	8	8	12	9	8		
Compuesta: Problemas escolares (en T2, T3, S2 y S3)		16	18				15	26
*Clínica: Problemas de atención		8	10					
*Clínica: Problemas de aprendizaje		8	8					
*Clínica: Actitud respecto a la escuela							8	8
*Clínica: Actitud respecto a los profesores							7	9
*Clínica: Búsqueda de sensaciones								9
Compuesta: Falta de atención/Hiperactividad (sólo en S2 y S3)							17	16
*Clínica: Hiperactividad							8	8
*Clínica: Problemas de atención							9	8
Compuesta: Índice de síntomas emocionales (sólo en S2 y S3)							52	64
*Adaptativa: Autoestima							7	7
*Adaptativa: Confianza en sí mismo							7	9
*Clínica: Ansiedad							11	13
*Clínica: Depresión							10	12
*Clínica: Estrés social							9	11
*Clínica: Sensación de ineptitud							8	12
Compuesta: Ajuste personal (sólo en S2 y S3)							32	36
*Adaptativa: Relación con los padres							10	11
*Adaptativa: Relaciones interpersonales							8	9
*Adaptativa: Autoestima							7	7
*Adaptativa: Confianza en sí mismo							7	9

- Para las escalas de contenido e índices (en cursiva, aparecen las escalas clínicas/adaptativas que forman parte de varias escalas compuestas):

<b>Escalas de contenido</b>	T1	T2	T3	P1	P2	P3	S2	S3
Acoso escolar	6	9	9	6	10	11		
Autocontrol emocional	14	12	10	17	10	12		
Control de la ira	7	7	7	8	9	8		10
Emocionalidad negativa	7	8	9	6	8	9		
Funcionamiento ejecutivo	17	24	27	23	20	22		
Resiliencia	6	13	12	9	9	13		
Trastornos del desarrollo social	16	15	15	17	19	18		
Ansiedad ante los exámenes								6
Fuerza del ego								8
Manía								10
<b>Índices clínicos</b>	T1	T2	T3	P1	P2	P3	S2	S3
Índice de deterioro funcional	10	44	44	29	44	41	20	28
Índice de probabilidad clínica	14			16				
Índice de probabilidad de autismo		22	15		13	18		
Índice de probabilidad de comportamiento perturbador		22	26		30	20		
Índice de probabilidad de TDAH		21	10		11	9		
<b>Índices de funcionamiento ejecutivo</b>	T1	T2	T3	P1	P2	P3	S2	S3
Índice de control atencional	5	8	10	7	7	9		
Índice de control conductual	5	7	6	6	6	7		
Índice de control emocional	8	7	6	7	4	4		
Índice de funcionamiento ejecutivo general	18	31	32	20	25	29		
Índice de resolución de problemas		9	10		8	9		

### 1.15. Soporte:

Papel y lápiz

Informatizado

### 1.16. Cualificación requerida para el uso del test de acuerdo con la documentación aportada:

Nivel C. Aplicación individual.

### 1.17. Descripción de las poblaciones a las que el test es aplicable:

Este cuestionario permite evaluar los problemas emocionales y conductuales de niños/as y adolescentes con edades comprendidas entre los 3 años y los 18 años y 11 meses, que sean susceptibles de presentar problemas emocionales y conductuales. Se recoge información de tres fuentes diferentes:

- Profesores/tutores (T). Este cuestionario está desarrollado para recoger información en el entorno educativo. Los profesores han de haber estado suficiente tiempo con el menor (diariamente durante un mes, o varios días a la semana durante 6 a 8 semanas). Si la evaluación se realiza durante el inicio de un curso escolar, se recomienda que el informante sea un profesor del curso anterior. Se señala que puede ser útil recoger información de varios informantes.

- Padres (P). Se recomienda, siempre que sea posible, que el cuestionario se aplique a ambos padres. Si tan sólo puede participar uno de los padres, es preferible que sea el que ha pasado más tiempo con el hijo/a. El manual no concreta si los padres han de tener un nivel educativo concreto, pero sí que recomienda revisar la información que se tiene de cada padre, y si muestran nerviosismo al contestar las escalas, porque puede haber problemas de lectura derivados de un bajo nivel educativo o de que su lengua materna sea otra. En ese caso, el profesional les ayudará a aclarar sus dudas.

- Menor (S, versión autoadministrada, con supervisión). El manual especifica que los niños con un bajo nivel educativo, con problemas de lectura, o si su lengua materna es otra, pueden generar respuestas inválidas por su bajo nivel de lectura. Recomiendan observar si muestran nerviosismo al contestar, o si requieren aclaraciones, porque puede ser un signo de problemas de lectura.

El manual no concreta si el test es aplicable en minorías étnicas o en personas con discapacidad. En el

manual se indica que la prueba permite evaluar el estado emocional y conductual de niños y adolescentes con déficits visuales y auditivos, siempre que el profesional que evalúa disponga de información y experiencia especializada en personas con este tipo de déficits. Respecto a los grupos clínicos, sí que se puede administrar en esos casos, ya que el cuestionario evalúa aspectos emocionales y problemas conductuales que son relevantes en estos grupos.

### 1.18. Existencia o no de diferentes formas del test y sus características:

Este cuestionario incluye subescalas específicas para los padres (P1, P2 y P3), escalas para los tutores (T1, T2 y T3) y escalas autoadministradas por el menor (S2 y S3). De esta forma es posible alcanzar una visión global del menor, considerando la información que proporciona el propio menor y también sus padres y tutores. Estas escalas no son idénticas, ya que no todos los ítems están incluidos en cada una de las escalas.

El cuestionario se puede administrar en papel o mediante la plataforma online (los ítems son los mismos en ambos casos). No se describen requisitos inusuales del hardware.

### 1.19. Procedimiento de corrección:

Automatizado por ordenador (plataforma de corrección on-line).

### 1.20. Puntuaciones:

\* Obtención de las puntuaciones directas:

La corrección de la prueba se lleva a cabo mediante una plataforma online (Qglobal). El manual no proporciona una explicación detallada sobre cómo calcular las puntuaciones, ya que estos pasos los hace la plataforma online (p.ej., no proporciona información sobre la inversión de los ítems ni sobre otros aspectos relacionados con la corrección). Sin embargo, el Anexo A incluye la lista de ítems que componen las escalas clínicas, las escalas adaptativas, las escalas de contenido y los índices. Las puntuaciones en las escalas compuestas se obtienen a partir de las puntuaciones en las escalas clínicas y adaptativas, pero no se explicita el modo exacto en el que se calculan.

\* Tipos de puntuaciones y clasificación:

La plataforma proporciona las puntuaciones en escala T (media 50 y desviación típica 10), además de los percentiles. Las puntuaciones T no son normalizadas, sino transformaciones lineales de las puntuaciones directas. Las puntuaciones se pueden obtener con respecto a baremos sin diferenciar por sexo o diferenciando por este. Para las escalas clínicas, adaptativas de contenido y compuestas, se obtiene una clasificación en cinco categorías. El rango de puntuaciones T se clasifica en 30 o inferior, 31-40, 41-59, 60-69 y 70 o superior. Las escalas en las que una puntuación mayor implica ausencia de problemas (p.ej., adaptativas) se clasifican en Clínicamente significativo, En riesgo, Medio, Alto y Muy alto. Las escalas en las que una puntuación mayor indica presencia de problemas (p.ej., clínicas) se clasifican en Muy bajo, Bajo, Medio, En riesgo y Clínicamente significativo.

Para los índices de funcionamiento ejecutivo se obtienen únicamente puntuaciones directas, que se clasifican en tres categorías ("No elevado", "Elevado" y "Muy elevado").

\* Validez de las respuestas:

En el manual se detallan los procedimientos para analizar la calidad de las respuestas. En primer lugar, se señala que el BASC-3 permite un máximo de dos ítems no puntuables (sin respuesta o con respuesta múltiple) por escala. Además se obtienen una serie de índices para detectar patrones anómalos:

Índice F. Mide la presencia de una valoración excesivamente negativa. En los cuestionarios P y T la puntuación directa se obtiene contabilizando el número de veces que el informante ha respondido "Casi siempre" a ítems que describen conductas negativas y "Nunca" a ítems que describen conductas positivas. Para el cuestionario S se obtiene contabilizando el número de veces que se ha respondido "Verdadero" o "Casi siempre" a ítems que describen conductas o actitudes negativas y "Falso" o "Nunca" a ítems que describen conductas o actitudes positivas. Se obtiene una clasificación en tres categorías ("Aceptable", "Cautela", "Suma cautela").

Índice L. Mide la tendencia del niño o adolescente a mostrar una imagen extremadamente positiva de sí mismo. Los ítems que componen esta escala son aseveraciones poco realistas (p.ej., "Siempre digo la verdad") o ligeramente autocríticas que la mayoría de las personas reconocería como propias (p.ej., "Tengo algunas malas costumbres"). Se obtiene una clasificación en tres categorías ("Aceptable", "Cautela", "Suma cautela").

Índice V. Es un indicador básico de la validez de las puntuaciones del cuestionario de autoinforme (i.e., presencia de respuestas inverosímiles). Se puntúa el número de veces que se responde positivamente (i.e., "Verdadero", "Frecuentemente" o "Casi siempre") a aseveraciones inverosímiles (p.ej., "Bebo 50 vasos de leche cada día"). Se obtiene una clasificación en tres categorías ("Aceptable", "Cautela", "Suma cautela").

Índice de consistencia. Indica la presencia de distintas respuestas en ítems muy similares (p.ej., ítems con correlaciones muy altas, positivas o negativas). La puntuación se obtiene como una suma de las diferencias absolutas entre respuestas a pares de ítems correlacionados. Se obtiene una clasificación en tres categorías ("Aceptable", "Cautela", "Suma cautela").

Índice de patrones de respuestas. Indica la presencia de respuestas idénticas en ítems consecutivos, o respuestas cíclicas. La puntuación es el número de veces que la respuesta a un ítem es distinta a la respuesta del ítem anterior. Se obtiene una clasificación en tres categorías ("Cautela-bajo", "Aceptable", "Cautela-alto").

### 1.21. Escalas utilizadas:

Percentiles

Puntuaciones estandarizadas T (Media 100 y desviación típica 50)

### 1.22. Posibilidad de obtener informes automatizados:

Sí. La plataforma online proporciona un informe a partir de las respuestas de los sujetos. Concretamente, proporciona un resumen de las puntuaciones obtenidas en los índices relacionados con los patrones de respuestas anómalos (por ejemplo, respuestas idénticas en ítems consecutivos), indicando si los valores son aceptables. También muestra un gráfico con las puntuaciones en las diferentes escalas simples, y una tabla con las puntuaciones T y percentiles en cada escala simple. A continuación, se muestra un resumen de las puntuaciones en las escalas compuestas, en las escalas clínicas y en las escalas adaptativas (puntuación T, percentil, intervalo de confianza, etc.). Se muestra después el gráfico del perfil de puntuaciones en las escalas de contenido y en los índices clínicos, y las tablas de puntuaciones de estas escalas e índices. Finalmente se presenta un listado con los ítems que pueden resultar de interés (estos ítems están resaltados en negrita), considerando las respuestas ofrecidas por los sujetos.

### 1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: Entre 10 y 20 minutos para las formas T y P. Entre 20 y 30 minutos para la forma S.

En aplicación colectiva: No procede [La aplicación sólo es individual].

### 1.24. Documentación aportada por el editor:

Manual. Cuestionarios.

### 1.25. Precio de un juego completo de la prueba:

Lapiz y papel [Manual, Cuestionario para padres (10 unidades P1, P2 y P3), Cuestionario para profesores (10 unidades T1, T2 y T3) y Cuestionario autoinforme (10 unidades S2 y S3) + 80 usos de corrección online, que pueden utilizarse indistintamente para cualquier cuestionario y forma]: 325€; Fecha de consulta: 16/02/2022.

Aplicación digital [Manual digital + 80 usos de corrección online, que pueden utilizarse indistintamente para cualquier cuestionario y forma]: 295€. Fecha de consulta: 16/02/2022.

### 1.26. Precio y número de ejemplares del paquete de cuadernillos:

42,50€ por cada tipo de cuestionario (25 ejemplares de ese cuestionario+25 correcciones online). Fecha de consulta: 16/02/2022.

### 1.27. Precio y número de ejemplares del paquete de hojas de respuesta:

No hay hojas de respuestas.

**1.28. Precio de la administración y/o corrección, y/o elaboración de informes por parte del editor:**

(está incluido en el precio del paquete de cuadernillos, ya que solamente puede corregirse a través de la plataforma online)

## 2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

### 2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Excelente	5
2.2. Calidad de la documentación aportada	Buena-Excelente	4.5
2.3. Fundamentación teórica	Buena	4
2.4. Adaptación del test	Adecuada - Buena	3.5
2.5. Desarrollo de los ítems del test	Buena	4
2.6. Calidad de las instrucciones para el participante	Excelente	5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Excelente	5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Buena	4
2.10. Datos sobre el análisis de los ítems	No se aporta	-

## 2.11. Validez:

### 2.11.1. Evidencias de validez de contenido:

Contenido	Valoración	Puntuación
2.11.1.1. Calidad de la representación del contenido o dominio	Buena	4
2.11.1.2. Consultas a expertos	Se ha consultado a un número moderado de expertos mediante un procedimiento sistematizado ( $10 \leq N \leq 30$ )	4

### 2.11.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

#### 2.11.2.1. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables:

Contenido	Valoración	Puntuación
2.11.2.1.1. Diseños empleados	Diferencias entre grupos, matriz multirasgo-multimétodo	
2.11.2.1.2. Tamaño de las muestras	Varios estudios con muestras pequeñas ( $N < 200$ )	2
2.11.2.1.3. Procedimiento de selección de las muestras	<p>Incidental. En la versión española de la BASC3, para diseñar una muestra lo más representativa posible de la población española, se siguieron las proporciones y los datos del Censo de Población y Viviendas español (INE, 2011). La muestra de tipificación se planeó sobre una muestra de 1075 evaluaciones de niños y adolescentes de la población general, estratificada según las variables demográficas edad, sexo y nivel educativo.</p> <p>Con respecto a la comparación de grupos, se utilizaron muestras de menores que debían cumplir los criterios diagnósticos del DSM-5 para el diagnóstico de TDAH o TEA, y que fueron seleccionadas en distintas zonas del país. Estas muestras son de tamaño reducido (<math>N</math> entre 8 y 54, según la versión del cuestionario).</p> <p>El manual presenta las correlaciones con otros tests obtenidas con la versión original del cuestionario, con una muestra de E.E.U.U., pero no para la adaptación española española, por lo que no se ha seleccionado la opción correspondiente en el</p>	



	ítem 2.11.2.1. Finalmente, en el estudio multirrasgo-multimétodo, se han llevado a cabo diferentes estudios con muestras de tamaño pequeño (máximo N = 178).	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	-	-
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	-	-
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	-	-
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	Adecuada	3
2.11.2.1.8. Resultados de las diferencias intergrupo	Buena	4

#### 2.11.2.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y un criterio:

Contenido	Valoración	Puntuación
2.11.2.2.1. Criterios empleados	No utilizan criterios externos para obtener evidencia de este tipo de validez. Como evidencia de validez se aportan los perfiles de puntuaciones de grupos de niños y adolescentes con determinados diagnósticos clínicos. Los resultados de los estudios relacionados con grupos especiales proporcionan pruebas de validez descritas en el apartado anterior.	
2.11.2.2.2. Calidad de los criterios empleados	-	-
2.11.2.2.3. Relación temporal entre test y criterio	-	
2.11.2.2.4. Tamaño de las muestras	-	-
2.11.2.2.5. Procedimiento de selección de las muestras	-	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	-	-

#### 2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Adecuado con carencias	2

<b>2.11.3.2. Funcionamiento diferencial de los ítems</b>	-	-
--	---	---

#### 2.11.4. Acomodaciones en la administración del test:

Contenido	Valoración	Puntuación
<b>2.11.4. El manual del test informa sobre las acomodaciones en la administración del test</b>	Sí. El manual ofrece recomendaciones para realizar una adecuada administración del cuestionario a personas con un bajo nivel educativo y con problemas de comprensión lectora.	

#### 2.11.5. Comentarios generales sobre evidencias de validez:

Las evidencias de validez se pueden resumir en los siguientes apartados:

- Evidencias de validez de contenido. El BASC-3 original, parte de una prueba original previa (el BASC). Para el desarrollo de la versión original se llevó a cabo un proceso de consulta a muestras amplias de profesores, padres y estudiantes para contemplar conductas positivas y negativas no recogidas en la versión previa. Por tanto, se ha intentado que haya una buena representación del dominio, siendo el número de escalas clínicas y adaptativas bastante amplio. En la adaptación de la prueba se consultó a un equipo de expertos en diferentes dominios de la psicología infanto-juvenil, que evaluó la traducción, valorando la idoneidad del contenido y la adecuación cultural de cada ítem. Aunque no se proporciona información cuantitativa y detallada de este proceso, se indica que se crearon algunos ítems nuevos para reemplazar aquellos que pudieran ser problemáticos.

- Evidencias de relaciones con otras variables. En el manual se presentan las correlaciones con otros cuestionarios para la versión original del cuestionario (muestra de Estados Unidos). Referente a la validez del cuestionario para profesores/tutores y del cuestionario para padres, las correlaciones entre las escalas del BASC3 y las escalas de otros cuestionarios (ASEBA, ConnerS3, D-Ref) que miden constructos similares son, por lo general, de moderadas a altas. En el caso del cuestionario autoinforme, los hallazgos son similares a los de los cuestionarios para profesores/tutores y para padres. Difieren en las medidas utilizadas para examinar la correlación con otros cuestionarios, que en este caso son el Conners 3, el RCMA2 y el BY-2. Considerando que en la adaptación española se han modificado diversos ítems, y en algunos casos la modificación afecta incluso al contenido de los ítems, estas correlaciones en la versión original no ofrecen evidencia sobre la validez del cuestionario adaptado al español. Por ese motivo, no se ha incluido nada sobre estas correlaciones en la valoración realizada en los ítems previos.

Por otra parte, se presentan las medias en diferentes subescalas para un grupo de menores con trastorno del espectro autista (TEA), un grupo con trastorno por déficit de atención/hiperactividad (TDAH). Se observa el patrón esperado. Para el grupo TDAH se esperan puntuaciones más altas que las de la población general en las escalas Problemas de atención e Hiperactividad. Para el grupo con TEA se esperan puntuaciones altas en varias escalas compuestas y clínicas (p.ej., Atipicidad), así como puntuaciones considerablemente por debajo del rango promedio en las escalas adaptativas. Convendría determinar qué diferencias son significativas, y cuál es la magnitud del efecto. En estos estudios, los tamaños muestrales para los grupos especiales son pequeños (i.e., para el cuestionario T, N < 20; para el cuestionario P, N < 55; para el cuestionario S, N < 48). Convendría incluir más estudios de este tipo considerando la variedad de conductas adaptativas y clínicas que se evalúan en el cuestionario.

En el manual se expone un estudio multirasgo-multimétodo, en el que se correlacionan las puntuaciones de las escalas equivalentes obtenidas a partir de diferentes informadores (el propio sujeto, padres/tutores o profesores), lo que constituiría una evidencia de validez. Entre cuestionarios P y T, las correlaciones de validez convergente para escalas clínicas, adaptativas y de contenido varían entre -.02 (Ansiedad, T1-P1) y .75 (Trastornos de desarrollo social, T1-P1), con un promedio de .42. Entre cuestionarios T y S, las correlaciones de validez convergente varían entre .13 (Ansiedad, T2-S2) y .47 (Hiperactividad, T2-S2), con un promedio de .30. Entre cuestionarios P y S, las correlaciones de validez convergente varían entre .05 (Atipicidad, P2-S2) y .53 (Problemas de atención, P3-S3), con un promedio de .33. El valor moderado de las correlaciones promedio es consistente con estudios previos, que muestran que las correlaciones de cuestionarios de autoinforme y otras medidas de la conducta infantil

son relativamente bajas. Esto puede interpretarse positivamente, indicando la necesidad de utilizar distintas fuentes de información, dado que pueden diferir las conductas percibidas u observadas en los distintos entornos (en el caso de la comparación entre los cuestionarios P y T) o puede haber diferencias entre informantes. A pesar de que los resultados son generalmente consistentes con estudios previos, existen varios casos en los que las correlaciones entre escalas que miden constructos distintos son similares o mayores que las correlaciones entre escalas que miden el mismo constructo, por lo que habría que profundizar en esos indicios de falta de validez discriminante. Las muestras para valorar la convergencia entre cuestionarios tienen tamaños muestrales entre 40 (p.ej., para T1, P1) y 178 (p.ej., para P3-S3).

Finalmente, el cuestionario tiene como objetivo evaluar problemas conductuales y emocionales en menores, pero debería recogerse más evidencia, en muestra española, sobre su utilidad como herramienta de apoyo para el diagnóstico. Aunque el cuestionario está orientado a ofrecer una imagen general de las debilidades, problemas y puntos fuertes del menor, y no al diagnóstico, sería importante incluir información que permita justificar los puntos de corte.

- Evidencias de validez sobre la estructura interna. Aunque las escalas primarias (clínicas y adaptativas) se han construido siguiendo criterios clínicos y de sensibilidad diagnóstica, con una fuerte validez de contenido, sería conveniente proporcionar evidencia, para la muestra española, sobre la composición/estructura de las escalas adaptativas y clínicas. Sí se realizan análisis factoriales para validar la composición de las escalas compuestas, cuya base de construcción original es el propio análisis factorial. La información aportada no es totalmente favorable a la composición de las escalas compuestas. Por ejemplo, el factor de interiorización no está definido en el cuestionario P1 y no se aporta evidencia de la separación del factor de Ajuste Personal y el factor de Problemas Escolares, en la escala S. Por otro lado, la presencia de pesos cruzados muy altos en la solución factorial (i.e., las escalas saturan casi siempre en más de un factor), refuerza la conveniencia de realizar análisis factoriales con relación a las escalas primarias, para identificar cuántos factores hay y qué ítems incluyen. También convendría que informaran del ajuste (utilizando métodos de estimación alternativos a ejes principales). Los análisis factoriales se han realizado en muestras cuyo tamaño varía entre 60 (para el cuestionario T1) y 198 (para el cuestionario P2).

Respecto a los índices de funcionamiento ejecutivo, se realizaron análisis factoriales de primer y de segundo orden. En alguna de las formas del BASC-3 obtuvieron valores de CFI y TLI inferiores a .90 y valores de RMSEA superiores a .07 (incluso de .10 o .11). Valores inferiores a .90 en CFI y TLI, y superiores a .07 en RMSEA son habitualmente considerados como inadecuados, lo que indicaría un mal ajuste.

- Análisis del funcionamiento diferencial de los ítems. Se informa de análisis del sesgo en la muestra de tipificación estadounidense, pero no en la muestra española, por lo que no se han considerado en esta evaluación.

## 2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	<p>Coefficientes de fiabilidad para diferentes grupos de personas.</p> <p>Error típico de medida para diferentes grupos de personas.</p>	

### 2.12.2. Equivalencia formas paralelas:

Contenido	Valoración	Puntuación
2.12.2.1. Tamaño de las muestras	-	-
2.12.2.2. Puesta a prueba de los supuestos de paralelismo	-	-
2.12.2.3. Promedio de coeficientes de equivalencia	-	-

### 2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Un estudio con una muestra de tamaño moderado ( $200 \leq N < 500$ ) para cada versión del cuestionario (P, T ó S)	4
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa o KR-20	
2.12.3.3. Promedio de coeficientes de consistencia	Excelente ( $r \geq 0.85$ )	5

### 2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Varios estudios con muestras pequeñas ( $N < 100$ )	2
2.12.4.2. Coeficientes de estabilidad	Excelente ( $r \geq 0.80$ )	5

### 2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	-	-
2.12.5.2. Coeficientes proporcionados	-	-
2.12.5.3. Tamaño de los coeficientes	-	-

### 2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	-	-
2.12.6.2. Promedio de los coeficientes	-	-

### 2.12.7. Comentarios generales sobre evidencias de fiabilidad:

Las evidencias de fiabilidad de las puntuaciones se resumen en los siguientes puntos:

1) Consistencia interna. Como este cuestionario tiene diversas versiones (en función de la edad del menor evaluado, y en función de si el informante es el propio menor, los padres o los profesores), el manual presenta los coeficientes alfa de Cronbach obtenidos a partir de cada una de estas versiones. Si se tiene en cuenta únicamente la diferenciación por versión, las muestras varían entre 60 (para T1) y 198 sujetos (para P2). Los valores obtenidos son buenos o excelentes. Sólo en unas pocas subescalas hay algún valor comprendido entre 0,60 y 0,70 (i.e., Ansiedad, en el cuestionario T1; Acoso Escolar, en los cuestionarios T1 y P1; Autoestima en S2; Locus de control en S3, grupo de edad 12-14).

2) Fiabilidad test-retest. Respecto a las correlaciones test-retest se aportan datos para la adaptación española de las versiones P y S (para la escala T se aportan los datos de la versión original americana). El promedio del intervalo de tiempo entre el test y el retest varía entre 18 y 28 días. El promedio de los coeficientes de estabilidad es superior a 0,80 y, en general, son buenos o excelentes (sin excepciones para los valores corregidos por restricción de rango). Respecto al tamaño de las muestras utilizadas en los estudios de estabilidad, se ha realizado un estudio para cada versión del cuestionario, en la que ha participado una muestra inferior a 100 participantes en cada caso (para el cuestionario P, N varía entre 33 y 63; para el cuestionario S, N varía entre 22 y 63).

No se han realizado estudios TRI ni se presentan datos de fiabilidad inter-jueces. Sería conveniente realizar un estudio de fiabilidad inter-jueces para las formas P y T, lo que permitiría valorar la concordancia entre informantes a la hora de responder estos cuestionarios.

## 2.13. Baremación e interpretación de las puntuaciones:

### 2.13.1. Interpretación normativa de las puntuaciones:

Contenido	Valoración	Puntuación
2.13.1.1. Calidad de las normas	Varios baremos dirigidos a diversos estratos poblacionales	4
2.13.1.2. Tamaño de las muestras	Una muestra de tamaño moderado ( $300 \leq N < 500$ ) para cada versión del cuestionario: P, T ó S)	3
2.13.1.3. Aplicación de estrategia de tipificación continua	No. Se estima la distribución suavizada con el método de las curvas de Johnson (que determina la distribución a partir de media, desviación típica, asimetría y curtosis).	
2.12.1.4. Procedimiento de selección de las muestras	Incidental. Para conseguir que la muestra fuera heterogénea, se ha realizado un muestreo estratificando, partiendo de los datos aportados por el INE en relación a la edad, el sexo y el nivel educativo de los padres. Los datos que aportan en relación a la configuración final de la muestra, en relación a los valores ideales dentro de los diferentes estratos, sugiere que es una muestra suficientemente representativa de la población objetivo, en cuanto a las variables de estratificación. Se cuidó también la variabilidad en cuanto a la zona geográfica.	
2.12.1.5. Actualización de baremos	Excelente (menos de 10 años)	5

### 2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	-	-
2.13.2.2. Procedimiento empleado para fijar los puntos de corte	-	
2.13.2.3. Procedimiento de obtención del acuerdo inter-jueces	-	
2.13.2.4. Valor del coeficiente de acuerdo inter-jueces	-	-

### 2.13.3. Comentarios generales sobre baremación e interpretación de las puntuaciones:

Como se ha comentado previamente, el BASC-3 consta de diferentes formas: cuestionarios para profesores/tutores (T), cuestionarios para padres (P) y cuestionarios autoinforme (S). Además, los cuestionarios varían en función de la edad del menor (hay versiones T y P para evaluar a niños de 3-5 años, versiones para 6-11 años y versiones para 12-18 años, mientras que en la forma S hay versiones para 6-11 años y 12-18 años).

La muestra de tipificación se divide en cinco grupos de edad para los cuestionarios T y P, y en tres grupos de edad para el cuestionario S. Para cada forma del cuestionario, hay baremos para diferentes grupos de edad, que pueden ser específicos por sexo o generales, sin diferenciar por sexo, ya que se observan diferencias en puntuaciones entre hombres y mujeres en algunas escalas (p.ej., los hombres obtenían puntuaciones más altas en las escalas Problemas de Atención, Hiperactividad y Búsqueda de sensaciones, mientras que las mujeres puntuaban más alto en Somatización). El uso de un baremo u otro puede depender del grado en que se valora que las diferencias entre las puntuaciones directas obtenidas por ambos sexos se deben simplemente a artefactos psicométricos o más bien indican que la presencia del problema (o la habilidad) es susceptible de estar más presente en un sexo que en el otro.

Para conseguir que la muestra utilizada para desarrollar estos baremos fuera heterogénea, se realizó un muestreo estratificando, partiendo de los datos aportados por el INE en relación con la edad, el sexo y el nivel educativo de los padres. También se cuidó la variabilidad en cuanto a la zona geográfica de procedencia. Por tanto, puede considerarse que las muestras son bastante representativas de la población objeto de estudio. Dado que la adaptación española del BASC-3 es reciente, se puede considerar que los baremos están actualizados. La muestra global es amplia (N = 1144), pero es moderada cuando se tiene en cuenta el tamaño muestral disponible para cada versión distinta del cuestionario: T (N = 342), P (N = 480) y S (N = 322) y tiende a ser más pequeña cuando se diferencia por versión para cada grupo de edad (T1, N = 60; T2, N = 162; T3, N = 120; P- 1, N = 90; P2, N = 198; P3, N = 192; S2, N = 130; S3, N = 192). Si además se tiene en cuenta que en los baremos se estratifica por sexo y edad, son excesivamente pequeñas (N < 66). Por ello en el manual se recomienda utilizar los baremos combinados por sexo, ya que en la mayoría de los casos permiten llevar a cabo un diagnóstico más preciso.

El manual indica los ítems que componen cada escala en un anexo. A través de la plataforma online puede obtenerse la puntuación directa, la puntuación T y el percentil en cada subescala e índice.

Respecto a los puntos de corte, proponen puntos de corte estadísticos (p.ej., en una escala clínica, puntuaciones T comprendidas entre 60 y 69 son interpretadas como "en riesgo", mientras que puntuaciones iguales o superiores a 70 son interpretadas como "clínicamente significativo"). En el manual explican que algunas de estas escalas tienen distribuciones asimétricas (es el caso de la escala depresión, por ejemplo, ya que la mayoría de los niños y adolescentes no presentan depresión), por lo que los autores recomiendan interpretar estas puntuaciones conjuntamente con los percentiles. Por último, aunque las escalas no tienen un propósito diagnóstico, sería recomendable una mayor justificación de los puntos de corte, que justifique su uso como una ayuda para este.