

Evaluación de

Sistema de cribado

conductual y emocional

 **BESS** 

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	BESS Sistema de cribado conductual y emocional (BESS del BASC-3)
Autor	Randy W. Kamphaus y Cecile R. Reynolds
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 niños y adolescentes, desde preescolar hasta bachillerato
Áreas de aplicación	Psicología clínica Psicología educativa Salud general y bienestar
Soporte	Papel y lápiz Informatizado

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Buena-Excelente	4.5
Fundamentación teórica	Buena	4
Adaptación	Buena	4
Análisis de ítems	--	--
Validez: contenido	Buena	4
Validez: relación con otras variables	Adecuada-Buena	3.9
Validez: estructura interna	Adecuada con algunas carencias	2
Validez: análisis del DIF	--	--
Fiabilidad: equivalencia	--	--
Fiabilidad: consistencia interna	Adecuada-Buena	3.3
Fiabilidad: estabilidad	Adecuada-Buena	3.5
Fiabilidad: TRI	--	--
Fiabilidad inter-jueces	Buena	4
Baremos e interpretación de puntuaciones	Buena	4

Comentarios generales

El BESS del BASC-3 (Sistema de evaluación de la conducta de niños y adolescentes – 3) es un instrumento diseñado para evaluar los puntos fuertes y débiles conductuales y emocionales en niños y adolescentes, desde preescolar hasta bachillerato. Se trata de un instrumento de cribado, por lo que pretende determinar el nivel de riesgo que presenta un niño/a o adolescente de desarrollar problemas conductuales y emocionales que requieran algún tipo de intervención. Es aplicable a distintos contextos de evaluación (educativo, clínico, de orientación psicológica...). Dispone de versiones para varios informantes: profesores/tutores (T), padres (P) y autoinforme (S), cuya longitud oscila entre 14 y 29 ítems dependiendo de la forma. En este sentido, es un test que puede ser atractivo por la brevedad de los cuestionarios, cuya administración requiere de alrededor de 5 minutos en aplicación individual y 15 en aplicación colectiva. Se valora positivamente que pueda administrarse tanto en lápiz y papel como de forma informatizada, aunque en ambos casos la corrección se realiza automáticamente a través de la plataforma Q-global. La administración y corrección a través de la plataforma informática contribuye a reducir posibles errores de corrección, aunque, al no proporcionar suficiente información para realizar correcciones manuales, se desconoce cómo se obtienen las puntuaciones directas (suma, suma ponderada, etc.). El manual proporciona información clara y detallada sobre las instrucciones de administración y ofrece recomendaciones sobre el proceso de evaluación. Además, los materiales son atractivos y están bien organizados para que el informante pueda rellenar los cuestionarios fácilmente.

El BESS proporciona una puntuación global: el Índice de riesgo conductual y emocional (IRCE). También cuenta con índices específicos, que representan aspectos más concretos del funcionamiento conductual y emocional del niño/a o adolescente. Adicionalmente, provee tres índices de validez que informan sobre la calidad de las respuestas emitidas. Ofrece un informe informatizado que presenta las puntuaciones obtenidas (directas y baremadas) y la interpretación del riesgo asociado, atendiendo a un sistema de clasificación (riesgo normal, elevado y sumamente elevado). El manual ofrece recomendaciones sobre la elección del baremo más apropiado según el contexto de evaluación y directrices claras de interpretación de los distintos índices proporcionados.

Las características sociodemográficas de la muestra con la que se analizan las propiedades psicométricas, aun habiendo utilizado un muestreo incidental, son similares a las del censo poblacional en cuanto a edad, sexo y nivel educativo de los padres. Además, se presta una atención especial a garantizar que se vean representadas diversas zonas geográficas de España.

El manual aporta diferentes fuentes de evidencias de validez. En lo relativo a las evidencias referidas al contenido, cabe señalar que el BESS consiste en una versión abreviada del BASC-3, cuyas propiedades psicométricas son satisfactorias, aunque en el manual del BESS, en varias ocasiones, se remite al lector a consultar el manual del BASC-3 para conocerlas. Este hecho dificulta la evaluación de la calidad del BESS teniendo en cuenta únicamente la información proporcionada en su manual. Un punto de mejora supondría documentar detalladamente las estrategias y procedimientos seguidos en el proceso adoptado en la selección, eliminación, sustitución y adición de ítems respecto al BASC-3 que permitan justificar la relevancia y representatividad de la composición final de ítems del BESS.

Con relación a las evidencias basadas en la relación con variables externas, se muestran los resultados de diversos estudios: (1) correlación entre el BESS y diversas medidas de constructos relacionados (evidencia convergente); (2) descripción (media y desviación típica) de las puntuaciones del BESS de dos grupos especiales (trastorno por déficit de atención con hiperactividad y trastorno del espectro autista) (evidencia basada en la comparación de grupos); y (3) análisis de la concordancia entre el BESS y el BASC-3 por lo que respecta a la clasificación del nivel de riesgo (evidencia basada en la relación con un criterio). Se valora positivamente la multiplicidad de estrategias empleadas, a pesar de que prácticamente todos los estudios dedicados a recabar evidencia convergente utilizan datos estadounidenses, limitando al caso español el estudio de la relación entre el BESS y el BASC-3.

Las evidencias de validez basadas en la estructura interna se analizan mediante la aplicación del análisis factorial confirmatorio y la correlación entre los índices del BESS. Si bien las cargas factoriales son, por lo general, superiores a 0.40, los índices de bondad de ajuste del análisis factorial muestran un ajuste al modelo marginal, por lo que la estructura dimensional de las diferentes versiones del BESS no se corresponde con la planteada a nivel teórico. Las correlaciones entre los índices del BESS presentan la dirección y magnitud esperadas según los valores que constan en el manual del BASC-3. Ampliar el número de evaluaciones para cada una de las versiones del BESS permitiría recabar evidencias más concluyentes sobre la estructura factorial y estudiar su invarianza para grupos relevantes, por ejemplo, definidos en función del sexo. De forma complementaria, sería recomendable estudiar el funcionamiento diferencial de los ítems del BESS.

La fiabilidad de las puntuaciones del BESS, analizada en el contexto de la Teoría Clásica de los Test, se considera satisfactoria, tanto en relación con la consistencia interna y la estabilidad de la medida, como con la concordancia entre informantes. También se ofrecen en una tabla los valores del error típico de medida, separados por grupo de edad y sexo, tanto para el IRCE como los índices específicos. De esta manera, puede obtenerse el intervalo de confianza de la puntuación obtenida.

Los baremos del BESS son actuales y adecuados a la población de interés, con tamaños muestrales entre suficientes (baremos específicos por sexo) y moderados (baremos combinados). Se basan en un total de 1144 evaluaciones, a saber, 342 para la forma P, 480 para el cuestionario T y 322 para la forma S, que se reducen a la mitad cuando son baremos separados por sexo.

En resumen, el BESS es una herramienta práctica, por su brevedad y facilidad de administración y corrección, y atractiva para los profesionales que desean identificar a niños y adolescentes en riesgo de desarrollar problemas conductuales y emocionales. Sus propiedades psicométricas son, por lo general, satisfactorias y avalan el uso del BESS en los diversos contextos de evaluación para los que ha sido diseñado. Convendría ampliar los estudios de validez de la versión española del BESS, especialmente en cuanto a su estructura factorial se refiere.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

BESS Sistema de cribado conductual y emocional (BESS del BASC-3)

1.2. Nombre del test en su versión original:

BASC-3 Behavioral and Emotional Screening System (BASC-3 BESS)

1.3. Autor/es del test original:

Randy W. Kamphaus y Cecile R. Reynolds

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:

2022

1.9. Fecha de la última revisión del test:

2022

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

Personalidad, Escalas clínicas, Problemas conductuales y emocionales

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

El test de cribado BESS pretende evaluar los problemas conductuales y emocionales de niños y adolescentes, con edades comprendidas entre los 3 años y los 18 años. Evalúa un amplio espectro de conductas y emociones, tales como la hiperactividad, la ansiedad, la comunicación funcional o los problemas de atención, entre otras. Se trata de un instrumento de cribado que pretende determinar el nivel de riesgo que presenta un niño/a o adolescente de desarrollar problemas conductuales y emocionales que requieran algún tipo de intervención. El cuestionario está diseñado para ser utilizado en centros educativos, centros de salud mental y orientación psicológica, consultorios de pediatría y centros de salud, y también en estudios de investigación.

Dispone de tres versiones según el informante -profesores/tutores (T), padres (P) y autoinforme (S)-, las cuales, a su vez, disponen de cuestionarios diferentes en función de la edad del niño/a o adolescente. Así, el BESS dispone de los siguientes cuestionarios:

- o Cuestionario para profesores/tutores (T-1): preescolar, de 3 a 5 años.
- o Cuestionario para profesores/tutores (T-2/T-3): niños/adolescentes, de 6 a 18 años.
- o Cuestionario para padres (P-1): preescolar, de 3 a 5 años.
- o Cuestionario para profesores/tutores (P-2/P-3): niños/adolescentes, de 6 a 18 años.
- o Cuestionario autoinforme (S-2/S-3): niños/adolescentes, de 8 a 18 años.

Para los diferentes cuestionarios, el BESS proporciona tres índices de validez que permiten valorar la calidad de los cuestionarios cumplimentados:

- o Índice F: permite evaluar la tendencia del informante a ser excesivamente negativo en sus valoraciones.
- o Índice de consistencia: permite detectar casos en los que la persona informante ha dado respuestas distintas a ítems que, por lo general, suelen recibir respuestas similares.
- o Índice de patrón de respuestas: permite identificar casos que pueden ser inválidos porque el informante ha respondido siguiendo un patrón concreto en vez de hacerlo atendiendo al contenido de los ítems.

Del mismo modo, para todos los cuestionarios, el BESS ofrece una puntuación global denominada Índice de riesgo conductual y emocional (IRCE) que indica el nivel de riesgo que presenta un niño/a o adolescente de tener o desarrollar problemas conductuales o emocionales.

Además de estos cuatro índices, disponibles para las versiones T, P y S, el BESS ofrece otros índices específicos que representan aspectos más concretos del funcionamiento conductual y emocional del niño/a o adolescente y que varían según la versión del BESS.

Las versiones T y P¹ disponen de los siguientes tres índices específicos:

- o Índice de riesgo de exteriorización: formado por ítems que reflejan conductas disruptivas (p.ej., agresividad y problemas de conducta).
- o Índice de riesgo de interiorización: formado por ítems asociados a conductas no disruptivas (p.ej., ansiedad y depresión).
- o Índice de riesgo de habilidades adaptativas: formado por ítems relacionados con conductas que reflejan habilidades adaptativas (p.ej., adaptabilidad, actividades cotidianas y comunicación funcional).

La versión S dispone de los siguientes tres índices específicos:

- o Índice de riesgo de interiorización: formado por ítems asociados a conductas que pueden indicar un cierto malestar psicológico (p.ej., atipicidad, locus de control y estrés social).
- o Índice de riesgo de autorregulación: formado por ítems asociados a conductas de autocontrol (p.ej., hiperactividad y problemas de atención).
- o Índice de riesgo de ajuste personal: formado por ítems relacionados con habilidades conductuales (p.ej., relación con los padres y relaciones interpersonales).

1.12. Áreas de aplicación:

¹ Las formas T-1 y P-1 no disponen del Índice de riesgo de interiorización.

1.13. Formato de los ítems:

Respuesta graduada / Tipo Likert

1.14. Número de ítems:

En la tabla siguiente se muestra el número de ítems para la obtención de los índices de validez en cada uno de los cuestionarios.

	T1	T-2/T-3	P1	P-2/P-3	S-2/S-3
Índices de validez	Número de ítems				
Índice F	13	9	11	13	10
Índice de patrón de respuestas	14	20	20	29	28
	Número de pares de ítems*				
Índice de consistencia	5	8	7	7	8

* Cada par contiene ítems que, por lo general, suelen recibir respuestas similares

Del mismo modo, en la tabla siguiente se muestra el número de ítems de los índices específicos y el IRCE, así como de cada uno de los cuestionarios en su conjunto. Las celdas sombreadas indican los casos para los cuales no existe el índice correspondiente (p.ej., el cuestionario T-1 no ofrece el Índice de riesgo de autorregulación).

	T-1	T-2/T-3	P-1	P-2/P-3	S-2/S-3
Índices específicos	Número de ítems				
Riesgo de exteriorización	6	6	8	9	
Riesgo de interiorización	1	6	3	9	10
Riesgo de habilidades adaptativas	5	5	6	9	
Riesgo de autorregulación					6
Riesgo de ajuste personal					8
	Número de ítems adicionales				
Riesgo conductual y emocional (IRCE)	2	3	3	2	4
Número total de ítems	14	20	20	29	28

1.15. Soporte:

Papel y lápiz, Informatizado

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

Nivel C

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

El BESS está diseñado para evaluar los puntos fuertes y débiles conductuales y emocionales de niños y adolescentes con edades comprendidas entre los 3 y los 18 años y 11 meses, desde preescolar hasta bachillerato. Permite recoger información de tres informantes diferentes:

- Profesores/tutores (T). Con este cuestionario se recoge información en el entorno educativo. Los profesores/tutores deben haber pasado un tiempo considerable con el niño/a o adolescente siendo suficiente que hayan estado con él/ella diariamente durante un mes o varios días a la semana durante, por lo menos, 6-8 semanas.
- Padres (P). Este cuestionario está diseñado para que lo cumplimente la madre, el padre o el tutor legal. Se recomienda que, en la medida de lo posible, se obtenga información de ambos padres; en caso de que únicamente pueda participar uno de los padres, es preferible que sea el que tiene más contacto con el niño/a o adolescente.
- Niño/a o adolescente (S). Se trata de un cuestionario autoadministrado, diseñado para que pueda ser cumplimentado a partir de los 8 años.

Según indica el manual, se puede utilizar en distintos entornos (p.ej., educativos, clínicos, de orientación psicológica, etc.) con el fin de identificar, de una forma rápida, potenciales problemas conductuales o emocionales de un niño/a o adolescente concreto o a nivel grupal. En este sentido, puede administrarse en centros educativos, centros de salud mental y orientación psicológica, consultorios de pediatría y estudios de investigación. En el manual se ofrecen directrices y recomendaciones acerca de la selección de los niños o adolescentes que formarán parte de la evaluación (p.ej., del último curso de educación infantil), la selección de los cuestionarios, la determinación de la frecuencia de las evaluaciones (p.ej., anualmente) y los informantes de los cuestionarios, en función del contexto de evaluación, los objetivos y la edad de los niños.

La muestra de tipificación constituye una muestra representativa de la población española, teniendo en cuenta la edad, el sexo y el nivel educativo de los padres como criterios de establecimiento de cuotas. Los niños y adolescentes de dicha muestra fueron sujetos hispanohablantes nativos y sin ningún tipo de discapacidad física o psíquica ni cualquier otra afección que interfiriera con la normal ejecución de la prueba.

Adicionalmente, se aportan datos del IRCE de dos grupos especiales, a) niños y adolescentes con edades comprendidas entre los 6 y los 18 años que cumplieran los criterios del DSM-5 para el diagnóstico del trastorno por déficit de atención/hiperactividad; y b) niños y adolescentes con edades comprendidas entre los 7 y los 18 años que cumplieran los criterios del DSM-5 para el diagnóstico del trastorno del espectro autista. En ambos casos un criterio de inclusión era haber obtenido una puntuación superior o igual a 80 en una prueba estandarizada de aptitud intelectual.

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

El BESS ofrece tres versiones según el informante: tutores/profesores (T), padres (P) y niños/adolescentes (S). Estas tres versiones, en combinación con la edad del niño/a o adolescente evaluado, resultan en los cinco cuestionarios que proporciona el BESS (ver punto 1.14): T1, T2-T3, P1, P2-P3, y S2-S3.

Los cinco cuestionarios pueden aplicarse mediante lápiz y papel o por internet. En ambos casos los cuestionarios son idénticos y únicamente cambia la forma de administración. El sistema de corrección por internet de Pearson (Q-global) permite tanto la aplicación como la corrección informatizada de los cuestionarios, además de la elaboración del informe de resultados, a nivel individual o grupal según corresponda. Si la aplicación se realiza en papel, se precisa introducir las respuestas en la plataforma para que el propio sistema pueda corregir el cuestionario y elaborar el informe correspondiente.

1.19. Procedimiento de corrección:

Automatizado por ordenador (plataforma de corrección on-line)

1.20. Puntuaciones:

El BESS ofrece diversas puntuaciones: una puntuación para cada uno de los tres índices de validez, una para el IRCE y una puntuación para cada uno de los índices específicos, los cuales varían en función del cuestionario que se haya aplicado.

Independientemente de que el BESS se haya aplicado en papel o de forma informatizada, la corrección de la prueba se realiza mediante la plataforma de corrección de test por internet Q-global. Es el propio sistema el que, según el baremo escogido (combinado por sexo o específico por sexo), calcula automáticamente las puntuaciones directas para los tres índices de validez (Índice F, Índice de patrón de respuestas, Índice de consistencia) y los índices específicos, así como las puntuaciones directas,

percentiles y T para el IRCE. Estas puntuaciones se proporcionan en el informe de resultados, junto con la valoración cualitativa correspondiente (p.ej., para el IRCE el nivel de riesgo presenta tres niveles: normal, elevado y sumamente elevado).

En cuanto al IRCE y los índices específicos, puntuaciones directas más elevadas son indicativas de mayor riesgo del niño/a o adolescente de presentar o desarrollar problemas conductuales o emocionales, con la excepción del Índice de riesgo de habilidades adaptativas y el Índice de riesgo de ajuste personal, donde puntuaciones más elevadas representan un riesgo menor.

En lo que se refiere a los índices de validez, puntuaciones más elevadas representan una menor calidad de las respuestas de los informantes. En el manual se proporcionan los detalles de la obtención de las puntuaciones directas de dichos índices, las cuales se presentan en el informe de resultados junto con la valoración cualitativa correspondiente (p.ej., para el Índice de consistencia, la validez de los datos presenta tres niveles: aceptable, cautela, y suma cautela).

1.21. Escalas utilizadas:

Centiles, T (Media 50 y desviación típica 10)

1.22. Posibilidad de obtener informes automatizados:

Sí. El sistema de corrección por internet Q-global proporciona un informe de resultados del BESS. Al inicio de dicho informe se presentan los datos identificativos del niño/a o adolescente e información relativa a la aplicación (p.ej., informante, baremo utilizado). En la parte central se proporcionan las puntuaciones de los índices de validez (puntuaciones directas y valoración del nivel de cautela), del IRCE (puntuaciones directas, percentiles y T y valoración del nivel de riesgo) y de los índices específicos (puntuaciones directas y valoración del nivel de riesgo). De forma adicional, para estos últimos, se presenta un listado de los ítems que contribuyen a los índices que se sitúan en un nivel de riesgo elevado o sumamente elevado. En la parte final del informe se incluye un apartado con las respuestas concretas que se han dado a cada uno de los ítems del IRCE.

Pueden verse ejemplos de informes ficticios en el apartado de Descargas de la web (<https://www.pearsonclinical.es/bess-sistema-de-cribado-conductual-y-emocional-del-basc-3>).

1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: 5 minutos.

En aplicación colectiva: 15 minutos.

1.24. Documentación aportada por el editor:

Manual, cuestionarios y clave de acceso para la aplicación y corrección por internet.

1.25. Precio de un juego completo de la prueba:

185 euros (IVA incluido) por un juego completo de aplicación manual: Manual, Cuestionario para padres (10 unidades P1, P2/P3), Cuestionario para profesores (10 unidades T1, T2/T3) y Cuestionario autoinforme (10 unidades S2/S3) + 50 usos de corrección online (también permite la aplicación online). Los usos pueden utilizarse indistintamente para cualquier cuestionario; 160 euros (IVA incluido) por un juego completo de aplicación digital: Manual digital + 50 usos de corrección online (permite la aplicación online). Los usos pueden ser utilizados indistintamente para cualquier cuestionario. Fecha de consulta: 27/03/2023.

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

34,95 euros (IVA incluido) por 25 cuestionarios + Recarga Q-global (25 usos de aplicación y corrección online) de alguno de los cuestionarios (P1, P2-P3, T1, T2-T3, S2-S3). Fecha de consulta: 27/03/2023.

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

No hay hojas de respuesta propiamente; las respuestas se registran en los cuestionarios y se introducen en la plataforma Q-global. Fecha de consulta: 27/03/2023.

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

34,95 euros (IVA incluido) por la modalidad manual: 25 cuestionarios + Recarga Q-global (25 usos de aplicación y corrección online) de alguno de los cuestionarios (P1, P2-P3, T1, T2-T3, S2-S3); Digital: 29,95 euros (IVA incluido) por la modalidad digital: recarga de 25 perfiles online de cualquiera de los cuestionarios (P1, P2-P3, T1, T2-T3, S2-S3). Fecha de consulta: 27/03/2023.

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	4
2.3. Fundamentación teórica	Buena	4
2.4. Adaptación del test	Buena	4
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	--	--

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	Correlaciones con otros test Diferencias entre grupos	
2.11.2.1.2. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado o con alguna muestra grande y otras pequeñas	4
2.11.2.1.3. Procedimiento de selección de las muestras	<p>Incidental. En el análisis de las correlaciones del BESS con otros test se utilizan tanto datos españoles como datos estadounidenses.</p> <p>Para diseñar una muestra lo más representativa posible de la población española, se tuvieron en cuenta los datos del Censo de Población y Viviendas español (INE, 2011). La muestra de tipificación se planteó sobre una muestra de 1075 evaluaciones de niños y adolescentes de la población general (400 evaluaciones para el cuestionario P, 400 evaluaciones para el cuestionario T y 275 evaluaciones para el cuestionario S), estratificada según las variables demográficas edad, sexo y nivel educativo. La zona geográfica no fue una variable de estratificación, pero se planificó que estuvieran representadas todas las comunidades autónomas determinando cuatro grandes zonas (norte, sur, este y centro).</p> <p>Se aplicaron los siguientes criterios de inclusión y exclusión:</p> <ul style="list-style-type: none"> o Niños y adolescentes con edades comprendidas entre los 3 años y los 18 años y 11 meses (T y P) y entre los 8 años y los 18 años y 11 	

	<p>meses (S).</p> <ul style="list-style-type: none"> o Sujetos hispanohablantes nativos. o Niños y adolescentes sin ningún tipo de discapacidad física o psíquica ni cualquier otra afección que interfiriera con la normal ejecución de la prueba. <p>Para el reclutamiento de participantes se contactó con profesionales con experiencia en evaluación infantil y juvenil gracias a la colaboración de los colegios oficiales de psicólogos y logopedas de todo España. Se seleccionaron un total de 53 profesionales procedentes de distintas zonas geográficas, los cuales participaron en un curso de formación online donde se expusieron los objetivos y requerimientos de la fase de tipificación y los procedimientos específicos de aplicación del test para homogeneizar el proceso de aplicación.</p> <p>Tras el reclutamiento, se alcanzó una muestra efectiva de 1144 aplicaciones (342 evaluaciones para el cuestionario P, 480 evaluaciones para el cuestionario T y 322 evaluaciones para el cuestionario S), distribuidas de forma muy similar a los porcentajes teóricos del censo.</p> <p>En el caso de los estudios correlacionales con datos estadounidenses, se utilizaron muestras de niños y adolescentes con tamaños muestrales que oscilaron entre 39 y 173 (media de 92.1 participantes), seleccionados mediante un muestreo incidental.</p> <p>En el estudio sobre las diferencias entre grupos, en el que se aportan las medias y desviaciones típicas del IRCE de dos grupos especiales (datos españoles), el muestreo fue incidental. Un equipo de examinadores españoles reclutó participantes, en distintas zonas de España, de dos grupos especiales:</p> <p>(a) niños y adolescentes con edades comprendidas entre los 6 y los 18 años que cumplieran los criterios del DSM-5 para el diagnóstico del trastorno por déficit de atención/hiperactividad; y</p> <p>(b) niños y adolescentes con edades comprendidas entre los 7 y los 18 años que cumplieran los criterios del DSM-5 para el diagnóstico del trastorno del espectro autista.</p> <p>En ambos casos, para poder formar parte del estudio, los participantes tenían que haber obtenido una puntuación igual o superior a 80 en una prueba estandarizada de aptitud intelectual.</p>	
<p>2.11.2.1.4. Calidad de los tests empleados como criterio o marcador</p>	<p>Buena</p>	<p>4</p>
<p>2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares</p>	<p>Excelente ($r \geq 0.70$)</p>	<p>5</p>

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	--	--
2.11.2.1.8. Resultados de las diferencias intergrupo	Adecuada	3

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	El criterio empleado es el test BASC-3.	
2.11.2.2.2. Calidad de los criterios empleados	Buena-Excelente	4.5
2.11.2.2.3. Relación temporal entre test y criterio	Concurrente	
2.11.2.2.4. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado o con alguna muestra grande y otras pequeñas	4
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental. Aunque no se reporta directamente en el manual del BESS, se deduce que el procedimiento de selección de la muestra de este estudio es el mismo que el del estudio de las correlaciones del BESS con el BASC-3, descrito en el apartado 2.11.2.1.3.	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Adecuada ($0.35 \leq r < 0.45$)	3

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Adecuada con carencias	2
2.11.3.2. Funcionamiento diferencial de los ítems	--	--

2.11.4. Acomodaciones en la administración del test:

Contenido	Valoración	Puntuación
2.11.4. El manual del test informa sobre las acomodaciones en la administración del test		No

2.11.5. Comentarios generales sobre evidencias de validez:

A efectos de la validación del BESS, en el manual se describen detalladamente las características de la muestra española empleada (muestra de tipificación descrita en el apartado 2.11.2.1.3). Adicionalmente, algunos de los estudios centrados en la validez emplean datos estadounidenses.

Evidencia basada en el contenido

El BESS se desarrolló al mismo tiempo que los cuestionarios T, P y S del BASC-3. Sus ítems se incluyeron en los cuestionarios de la fase de tipificación del BASC-3, contando con alrededor de 775 ítems (unos 450 en los cuestionarios T y P y 325 en el S). En el manual se indica que, inicialmente, se seleccionaron aquellos ítems de la versión de tipificación del BASC-3 que, cubriendo las áreas específicas susceptibles de ser problemáticas a nivel conductual y emocional (puntuaciones compuestas), estaban libres de sesgo. Posteriormente, aquellos que presentaban las cargas más elevadas en un análisis de componentes principales y aportaban contenido exclusivo (no representado por ningún otro ítem) y mostraban unas propiedades psicométricas similares en las distintas formas (T y P). Finalmente, se seleccionó un número más reducido de ítems atendiendo al equilibrio en la representación de las áreas relevantes de las escalas compuestas y la facilidad de cumplimentación, y se añadieron algunos otros atendiendo a la importancia de su contenido, aunque no se aportan datos cuantitativos (p. ej., del grado de relevancia y representatividad de los ítems). En la traducción/adaptación del inglés al español, se contó con la colaboración de 10 expertos en psicología infanto-juvenil, entre los que se encuentra Randy W. Kamphaus, coautor de la versión original del BASC-3 y del BESS, miembros del equipo de Pearson Clinical & Talent Assessment, así como profesionales especialistas en infancia y adolescencia y también en traducción y corrección lingüística. Después de la traducción y adaptación de los ítems al español, se realizó un estudio piloto del conjunto de los ítems con una muestra de 10 niños y adolescentes para identificar posibles problemas de comprensión. Estas estrategias, en conjunto, aportan evidencias de validez basadas en el contenido del test.

Evidencias basadas en la relación entre las puntuaciones del test y otras variables

Se obtienen diferentes fuentes de validez a través de varios estudios:

(1) Evidencia convergente. Se correlacionan las puntuaciones del BESS con diversas medidas de ajuste conductual y emocional. Con datos españoles y en muestras entre pequeñas y moderadas, se obtienen las correlaciones de los cuestionarios del BESS con las del BASC-3 y se obtiene un valor promedio de 0.82 cuando se consideran los índices del BESS y del BASC-3 que evalúan aspectos similares. Este valor tan elevado puede venir explicado, al menos en parte, por el hecho de que ambos instrumentos comparten un número considerable de ítems. Con datos estadounidenses, las correlaciones se proporcionan entre los cuestionarios T, P y S y el *Achenbach System of Empirically Based Assessment* (ASEBA), las formas T-2/T-3, P-2/P-3 y S-2/S-3 y el *Conners 3*, los cuestionarios T y P y la *Autism Spectrum Rating Scales* (ASRS), y, finalmente, las formas S-2/S-3 y la *Revised Children's Manifest Anxiety Scale* (RCMAS-2). Estos últimos estudios se basan en muestras pequeñas, a saber, de entre 39 y 173 evaluaciones (media de 92.1), dependiendo de la forma del BESS y el instrumento utilizado como marcador. En los casos en que ambos instrumentos evaluaban aspectos similares, la mayoría de los valores de los coeficientes de correlación fueron de una magnitud superior a 0.60 (buena), llegando incluso a valores de más de 0.80 (excelente).

(2) Evidencia basada en la comparación entre grupos. El manual proporciona los descriptivos (media y desviación típica) de la puntuación del IRCE del BESS para dos grupos especiales (trastorno por déficit de atención con hiperactividad y trastorno del espectro autista). Aunque no se comparan estadísticamente las puntuaciones debido al reducido tamaño muestral (entre 8 y 54 dependiendo del cuestionario), como se esperaba, se observan puntuaciones más elevadas en los grupos especiales. Además de ampliar los

tamaños muestrales, sería recomendable contrastar estadísticamente las puntuaciones y acompañar dichos contrastes de la correspondiente medida del tamaño del efecto.

(3) Evidencia basada en la relación con un criterio. Se analiza la concordancia entre el BESS y el BASC-3 por lo que respecta a la clasificación del nivel de riesgo, proporcionando los valores de sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo. Al tratarse de un test de cribado, al seleccionar los puntos de corte más apropiados, los autores optaron por maximizar la sensibilidad y el VPP, manteniendo unos valores aceptables de especificidad y homogeneizando los puntos de corte entre cuestionarios del BESS. En promedio, el valor de la sensibilidad de los diferentes índices del BESS fue de 0.62, y de 0.73 para el IRCE. Estos valores indican que la capacidad del BESS para identificar a niños y adolescentes con problemas conductuales y emocionales es adecuada.

Evidencias basadas en la estructura interna

La estructura interna del BESS se estudia, en la muestra de tipificación, mediante un análisis factorial confirmatorio, con los índices específicos como factores de primer orden y el IRCE como factor de segundo orden. Sin conocer cuál es el método de estimación empleado, los valores de los índices de ajuste varían en función de la forma, con valores de CFI que oscilan entre 0.77 (T-1) y 0.86 (T-2/T-3) y de RMSEA entre 0.06 (S-2/S-3) y 0.12 (T-1). En este sentido, el ajuste del modelo fue marginal en todos los casos, por lo que la estructura dimensional de las diferentes versiones del BESS no se corresponde con la planteada a nivel teórico. Las cargas factoriales fueron, por lo general, superiores a 0.40, con la excepción de algunos ítems. Como consecuencia de las bajas cargas factoriales obtenidas en los ítems del Índice de interiorización de las formas T-1 y P-1, este índice no se obtiene en dichas formas. En el manual también se reportan las correlaciones entre los distintos índices que ofrece el BESS, obteniendo valores en la dirección y magnitud esperadas si se comparan con los obtenidos en la validación del BASC-3.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad (para cada escala o subescala) Coeficientes de fiabilidad para diferentes grupos de personas Error típico de medida para diferentes grupos de personas	

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 moderada ($200 \leq N < 500$) o varios estudios con muestras pequeñas ($N < 200$)	2
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa o KR-20	
2.12.3.3. Promedio de coeficientes de consistencia	Buena ($0.80 \leq r < 0.85$)- Excelente ($r \geq 0.85$)	4.5

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Un estudio con una muestra moderada ($100 \leq N < 200$) o 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	Coeficiente de correlación de Pearson	
2.12.6.2. Promedio de los coeficientes	Bueno ($0.70 \leq r < 0.80$)	4

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

El estudio de la fiabilidad de las puntuaciones del BESS se centra en la consistencia interna (muestras españolas), la estabilidad temporal (muestras españolas y estadounidenses) y la concordancia entre informantes (muestras estadounidenses).

En el estudio de la consistencia interna, se calculan los coeficientes de fiabilidad separadamente en función del grupo de edad (Edades: 3-5, 6-7, 8-11, 12-14 y 15-18) y el género (mujeres, varones y combinado). Los tamaños muestrales de los diferentes grupos de edad oscilan entre 60 y 130 niños y adolescentes. La consistencia interna se evalúa a través de dos estrategias diferentes. Por un lado, se aplica el procedimiento de las dos mitades para el IRCE. Su utilización queda justificada porque este índice incluye ítems que evalúan problemas conductuales y emocionales muy diversos. Se repartieron los ítems con contenido común entre dos partes y, posteriormente, se calculó el coeficiente de correlación de las dos mitades y se aplicó la fórmula de Spearman-Brown. Los resultados muestran coeficientes de fiabilidad del IRCE excelentes, con valores superiores a 0.85. Por otro lado, se obtiene el coeficiente alfa de Cronbach para los índices específicos, dado que los ítems que pertenecen a los correspondientes índices son más homogéneos. Con los datos combinados (mujeres y varones), se obtienen valores superiores a 0.70 para todos los índices. Se concluye que la consistencia interna de los índices del BESS es satisfactoria.

Con estas mismas muestras, se reporta el error típico de medida (ETM) en puntuaciones T para el IRCE y en puntuaciones directas para los índices específicos, el cual permite calcular el intervalo en el que se situaría la puntuación verdadera con un nivel de confianza determinado.

En el estudio de la estabilidad temporal de las puntuaciones del BESS, se presentan datos españoles para los cuestionarios P y S, y estadounidenses para la forma T. Los tamaños muestrales oscilan entre 33 (cuestionario P-1) y 177 (cuestionario T-2/T-3). El intervalo transcurrido entre las dos administraciones oscila entre 7 y 70 días para el cuestionario T, entre 7 y 90 para la forma P, y entre 5 y 90 días para el cuestionario S. Se presentan coeficientes, tanto sin corregir como corregidos por la variabilidad de la muestra de tipificación, teniendo en cuenta los baremos combinados por sexo. Los valores corregidos oscilan entre 0.82 y 0.99, por lo que se concluye que la estabilidad temporal de los índices del BESS es muy satisfactoria.

Finalmente, el estudio del acuerdo entre informantes (p.ej., dos profesores o el padre y la madre que por separado cumplimentan, para un mismo niño/a o adolescente, el

correspondiente cuestionario), se basa en datos estadounidenses para las formas T (tutores) y P (padres). Los tamaños muestrales oscilan entre 58 para el cuestionario T-1 y 231 para el cuestionario P-2/P-3 y el intervalo temporal medio entre las dos aplicaciones es de aproximadamente 11 días. Los coeficientes de acuerdo entre informantes oscilan entre 0.52 y 0.85 (Media= 0.73), lo que denota que la concordancia es buena.

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	Moderado ($300 \leq N < 600$)	3
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	<p>Incidental. Los baremos españoles del BESS se construyeron a partir de los datos de la muestra de tipificación. El procedimiento de selección de los participantes de esta muestra se encuentra descrito en el apartado 2.11.2.1.3.</p> <p>En los baremos combinados, se cuenta con una muestra de 342 evaluaciones para el cuestionario P, 480 evaluaciones para el cuestionario T y 322 evaluaciones para el cuestionario S. En los baremos por sexo (varones/mujeres), los tamaños muestrales son, respectivamente, de 171, 240 y 161.</p> <p>Además de estos baremos, a través de la plataforma Q-global, los profesionales pueden solicitar que las puntuaciones se comparen con las de un grupo proveniente de poblaciones específicas (diagnóstico del trastorno por déficit de atención/hiperactividad o del trastorno del espectro autista), cuyo procedimiento de selección se describe también en el apartado 2.11.2.1.3.</p>	
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	--	--

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

A partir de 1144 evaluaciones, se proporcionan 9 baremos diferentes, resultantes de cruzar los cuestionarios (T, P y S) con el sexo (mujer, varón y combinado). En la composición de la mayoría de las muestras se ha procurado que queden representadas diferentes características sociodemográficas relevantes (edad, nivel educativo de los padres, sexo y zona geográfica).

Los baremos ofrecidos a través de la plataforma de corrección online Q-global transforman las puntuaciones directas en puntuaciones T (media 50 y desviación estándar 10) y percentiles en el caso del IRCE. Para este índice también se proporcionan niveles de clasificación según la puntuación T obtenida: riesgo normal (0-60), riesgo elevado (61-70) y riesgo sumamente elevado (≥ 71). En el caso de los índices específicos, se ofrecen las puntuaciones directas y los correspondientes niveles de clasificación del riesgo, definidos a partir de la distancia en desviaciones típicas de la puntuación directa respecto a la media (una desviación típica, entre una y dos desviaciones típicas o más de dos desviaciones típicas). Las etiquetas otorgadas al nivel de riesgo coinciden con las del IRCE (normal, elevado y sumamente elevado).

Para cada cuestionario, el BESS ofrece dos opciones para interpretar las puntuaciones de los diferentes índices: baremos combinados por sexo y baremos específicos por sexo. En el manual se indica que la elección depende del objetivo de la evaluación y del tipo de entidad que la lleva a cabo y se ofrecen algunas recomendaciones al respecto. En los baremos combinados, se cuenta con una muestra de 342 evaluaciones para el cuestionario P, 480 evaluaciones para el cuestionario T y 322 evaluaciones para el cuestionario S. En los baremos por sexo (varones/mujeres), los tamaños muestrales son, respectivamente, de 171, 240 y 161. En este sentido, los tamaños muestrales oscilan entre suficientes (baremos específicos por sexo) y moderados (baremos combinados).

Además de estos baremos, a través de la plataforma Q-global, los profesionales pueden solicitar que las puntuaciones se comparen con las de un grupo proveniente de poblaciones específicas (diagnóstico del trastorno por déficit de atención/hiperactividad o del trastorno del espectro autista).