Consejo General de la Psicología
ESPAÑA

Miembro de la Federación Europea de Asociaciones de Psicólogos

Conde Peñalver, 45. 5º Izq.
28006 Madrid – España
Tel.: +34 91 444 90 20
Fax: +34 91 309 56 15
secop@cop.es
www.cop.es

Evaluación de

Inventarios de Beck para niños y adolescentes -2

BYI-2

Prueba valorada en la ronda de evaluaciones del año 2018

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Inventarios de Beck para niños y adolescentes - 2 (BYI-2)
Autor	Judith S. Beck, Aaron T. Beck, John B. Jolly y Robert A. Steer
Autor de la adaptación española	Dpto. I+D de Pearson Clinical & Talent Assessment: Ana Hernández, Cristina Aguilar, Èrica Paradell y Fréderique Vallar
Variable(s)	Depresión, ansiedad, ira, conducta disruptiva y autoconcepto
Áreas de aplicación	Psicología clínica, psicología educativa, psicología forense, servicios sociales, salud general y bienestar
Soporte	Papel y lápiz, informatizada. Corrección por internet

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Buena	4
Fundamentación teórica	Adecuada-Buena	3,5
Adaptación	Buena	4
Análisis de ítems	No se aporta información	---
Validez: contenido	Adecuada	3
Validez: relación con otras variables	Buena	4
Validez: estructura interna	Adecuada, pero con algunas carencias	2
Validez: análisis del DIF	No se aporta información	---
Fiabilidad: equivalencia	No se aporta información	---
Fiabilidad: consistencia interna	Buena-Excelente	4,5
Fiabilidad: estabilidad	Buena-Excelente	4,5
Fiabilidad: TRI	No se aporta información	---
Fiabilidad inter-jueces	No se aporta información	---
Baremos e interpretación de puntuaciones	Excelente	5

Comentarios generales

Los Inventarios de Beck para niños y adolescentes-2 (BYI-2) son cinco cuestionarios publicados por Pearson Educación, dirigidos a la evaluación de la depresión, la ansiedad, la ira, la conducta disruptiva y el autoconcepto de niños y adolescentes de edades comprendidas entre 7 y 18 años y 11 meses. Se trata de cuestionarios breves, autoaplicados, que pueden utilizarse por separado o combinados, de sencilla y rápida administración. Cada inventario consta de 20 afirmaciones, formuladas en primera persona, sobre pensamientos, sentimientos o conductas relacionadas con trastornos emocionales y sociales, que deben responderse con una escala de frecuencia de cuatro opciones (nunca, a veces, a menudo, siempre). El tiempo de aplicación ronda entre los 5 y 10 minutos para cada inventario.

El BYI-2 es la adaptación española de los estadounidenses Beck Youth Inventories-Second Edition (Beck, Beck, Jolly y Steer, 2005), desarrollados por la prevalencia observada de problemas de salud mental en niños y adolescentes, así como por la necesidad de evaluar el malestar psicológico de estos con rapidez y eficacia. Anteriormente se han utilizado ampliamente las escalas de Beck para adultos de depresión (BDI-II; Beck, Steer y Brown, 1996, 2011) y ansiedad (BAI; Beck y Steer, 1993, 2011) que, si bien cuentan con evidencias de utilidad clínica en adolescentes, requieren un nivel de comprensión lectora al menos de 5º de primaria y recogen aspectos que obtienen respuestas incongruentes (por ejemplo, ítems referidos a la sexualidad). El BYI-2, en comparación con las versiones para adultos, incluye contenidos más relevantes para la vida de los adolescentes (por ejemplo, sobre la escuela) y evalúa la ira y los síntomas externalizados (además de la depresión, la ansiedad y la autoestima). En este sentido, podría observarse cierto solapamiento entre los inventarios en el caso de los adolescentes con un nivel lector de 5º de primaria o superior, por lo que sería recomendable incluir criterios específicos que permitan recomendar la aplicación de una u otra prueba para este subgrupo.

Uno de los aspectos más destacables de los inventarios es que se dirigen a evaluar variables de gran importancia en el contexto de los servicios educativos, sanitarios y sociales. Los inventarios BYI-2 pueden ser de mucha utilidad para la detección de posibles situaciones de malestar psicológico entre niños y adolescentes. Resulta de gran interés, en la justificación teórica, la revisión realizada de los instrumentos existentes para evaluar la depresión, la ansiedad, la conducta disruptiva, la ira y el autoconcepto en niños y adolescentes, subrayando las limitaciones relacionadas con sus evidencias de fiabilidad y validez a partir de los criterios sugeridos por Prieto y Muñiz (2000). En este sentido, sería deseable incluir en la justificación teórica una explicación detallada de los modelos teóricos en los que se basan los inventarios, así como incluir referencias bibliográficas más actualizadas.

Otro aspecto destacable de los inventarios BYI-2 es que son de fácil y rápida aplicación, lo que resulta especialmente importante en la evaluación de niños y adolescentes. Los inventarios están formados por un número menor de ítems que la mayoría de las pruebas existentes, reduciendo considerablemente el tiempo de administración. Resulta una ventaja también el hecho de que puedan aplicarse de forma combinada o independiente, dando lugar a un amplio rango de posibilidades de evaluación en función de distintas necesidades u objetivos específicos. Por otra parte, pueden ser aplicados de forma individual y colectiva. Así, puede ser muy útil su aplicación grupal en contextos escolares para realizar cribados del posible malestar psicológico entre el alumnado. Se trata, por tanto, de un instrumento de gran versatilidad.

Los baremos son también destacables como punto fuerte del instrumento: basados en una muestra aleatoria de niños y adolescentes de población general, estratificada por sexo y edad. Para diseñar una muestra lo más representativa posible se siguieron las proporciones y datos del Censo de Población y Viviendas español (INE, 2011), ajustándose los porcentajes de la muestra de tipificación de forma satisfactoria a los porcentajes teóricos del censo. Se incluyen baremos por rangos de edad (7-10 años; 11-14 años, 15-18 años) y sexo (varones, mujeres).

Destaca también de forma positiva la corrección e interpretación de las pruebas. Las puntuaciones directas se transforman en una puntuación T para cada escala, lo que facilita el análisis del perfil para determinar hasta qué punto cada variable puede subyacer al malestar psicológico de un niño o un adolescente. Adicionalmente, en el manual se incluyen cuatro estudios de caso para ilustrar el proceso de interpretación. La corrección es sencilla y minimiza los errores al realizarse de forma informatizada.

En cuanto a las evidencias de fiabilidad de las puntuaciones (i.e., consistencia interna, error típico de medida, consistencia temporal), se obtienen valores satisfactorios en distintos estudios. Aunque la muestra total es de tamaño grande (N=703), sería interesante incrementar el número de participantes de lo subgrupos estudiados (por sexo y edad) e incluir otros coeficientes de consistencia interna más recomendados para datos ordinales (por ejemplo, el alfa ordinal).

Otro aspecto mejorable del manual está relacionado con la recomendación de incluir evidencias de validez basadas en el contenido de los inventarios. Sería recomendable incluir información específica que permitiera valorar la calidad de la representación de cada dominio, así como sobre el procedimiento de adaptación de los ítems, la consulta a expertos, la modificación e inclusión de nuevos ítems en la fase piloto y la eliminación de ítems tras la fase experimental.

También sería deseable proporcionar más evidencias de validez basadas en la estructura interna de la escala. En primer lugar, sería conveniente incluir más información acerca de los resultados del análisis de componentes principales (por ejemplo, cargas factoriales, valores propios, gráfico de sedimentación). En segundo lugar, en futuras ediciones del test sería conveniente, dado que se trata de unos inventarios derivados de un modelo conocido, lo realizar un análisis factorial confirmatorio (o modelos exploratorios de ecuaciones estructurales).

Se presentan, en cambio numerosas evidencias basadas en la relación con otras variables que son valoradas como satisfactorias y adecuadas, si bien en futuras ediciones del test cabría, por una parte, realizar una mejor justificación de la elección de las variables, su relación con los constructos medidos por los inventarios y los instrumentos usados para su evaluación. Por otro lado, sería deseable incluir estudios con muestras españolas y aumentar el tamaño de las muestras. Aun más más positivamente se valoran las evidencias basadas en un criterio externo proporcionadas, pues se aportan datos específicos de estudios realizados con muestras clínicas de niños y adolescentes españoles que cumplían criterios diagnósticos para uno o más trastornos (depresivos, de ansiedad, disruptivos, del control de impulsos y la conducta, alimentarios y de la ingestión de alimentos). Tales resultan de gran interés para realizar comparaciones y estudiar las diferencias con las puntuaciones obtenidas por un grupo control. Se facilita así el uso de los inventarios especialmente en el ámbito clínico y se aportan evidencias de su capacidad discriminativa.

Otros aspectos que podrían incluirse en ediciones posteriores son el análisis de DIF (por género y edad), especificar si se siguieron las directrices de la International Test Commission (2017) en la traducción y adaptación, y detallar los resultados del análisis de TRI.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Inventarios de Beck para niños y adolescentes - 2 (BYI-2).

1.2. Nombre del test en su versión original:

Beck Youth Inventories - Second edition.

1.3. Autor/es del test original:

Judith S. Beck, Aaron T. Beck, John B. Jolly y Robert A. Steer.

1.4. Autor/es de la adaptación española:

Departamento de I+D de Pearson Clinical & Talent Assessment: Ana Hernández, Cristina Aguilar, Èrica Paradell y Fréderique 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:

2005.

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

2017.

1.9. Fecha de la última revisión del test:

2017.

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

Escalas clínicas.

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

El Inventario de Beck para niños y adolescentes (BYI-2) es un instrumento dirigido a evaluar el malestar psicológico en niños y adolescentes con edades comprendidas entre 7 y 18 años y 11 meses (7:0-18:11). Más concretamente, consta de cinco inventarios autoaplicados, que se pueden utilizar por separado o combinados, centrados en la evaluación de la depresión, la ansiedad, la ira, la conducta disruptiva y el autoconcepto.

Cada inventario está formado por 20 afirmaciones sobre pensamientos, sentimientos o conductas relacionados con trastornos emocionales y sociales en niños y adolescentes. El formato de respuesta es una escala de frecuencia con cuatro opciones (nunca, a veces, a menudo, siempre). Los niños y adolescentes han de responder con qué frecuencia se identifican con cada una de dichas afirmaciones, incluyendo el propio día de la evaluación.

- El inventario de depresión de Beck para niños y adolescentes (BDI-Y) evalúa los pensamientos negativos del niño o del adolescente sobre sí mismo, su vida y su futuro; los sentimientos de tristeza e indicaciones fisiológicas de depresión. Tiene como objetivo identificar los síntomas de depresión de acuerdo con los criterios incluidos en los manuales diagnósticos.

- El inventario de ansiedad de Beck para niños y adolescentes (BAI-Y) evalúa los miedos y preocupaciones del niño o del adolescente (p. ej., sobre la escuela, hacerse daño, salud), así como síntomas físicos asociados a la ansiedad.

- El inventario de ira de Beck para niños y adolescentes (BANI-Y) refleja la percepción de recibir un trato injusto, pensamientos negativos sobre los demás, sentimientos de ira y activación fisiológica.

- El inventario de conducta disruptiva de Beck para niños y adolescentes (BDBI-Y) evalúa conductas y actitudes asociadas con trastornos disruptivos, el control de los impulsos y de la conducta.

- El inventario de autoconcepto de Beck para niños y adolescentes (BSCI-Y) mide la percepción que tienen de sí mismos en relación con competencias, fortaleza y autoestima.

Se trata de escalas breves y fáciles de aplicar (el tiempo de aplicación ronda entre los cinco y 10 minutos cada inventario), siendo posible tanto la aplicación individual como colectiva.

Puede aplicarse en distintos contextos para identificar a niños y adolescentes en una situación de riesgo (i.e., escolar; forense, residencial y hospitalario; ambulatorio de salud mental; atención primaria). Además, permite comparar las puntuaciones obtenidas por una persona con la de sus pares en edad o comparar la gravedad de los síntomas pasados y actuales. Como cualquier herramienta de evaluación, no debe utilizarse como medida única para establecer un diagnóstico.

1.12. Áreas de aplicación:

Psicología clínica, psicología educativa, psicología forense, servicios sociales, salud general y bienestar.

1.13. Formato de los ítems:

Respuesta graduada / tipo Likert.

1.14. Número de ítems:

100 ítems en total, 20 ítems en cada uno de los cinco inventarios:

BDI-Y= 20 ítems.

BAI-Y= 20 ítems.

BANI-Y= 20 ítems.

BDBI-Y= 20 ítems.

BSCI-Y= 20 ítems.

1.15. Soporte:

Papel y lápiz, informatizado. Corrección por internet.

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

Nivel C.

Si bien el BYI-2 puede ser aplicado por profesionales sin experiencia en el uso de los tests, estos deben estar siempre supervisados por una persona experta que se responsabilice de la aplicación e interpretación de los resultados. Esta persona experta debe estar familiarizada con los procedimientos de evaluación clínica, así como conocer los usos apropiados y las limitaciones de los tests psicológicos relacionadas con su validez y fiabilidad.

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

La prueba se puede administrar a niños y adolescentes de edades comprendidas entre 7 años y 18 años y 11 meses.

No se aporta información sobre su posible aplicación a poblaciones específicas o minoritarias si bien, dado que se trata de un instrumento autoinformado, requiere ciertas destrezas lectoras, de comprensión, atencionales e intelectuales. Por ello se especifica que, si la persona evaluada lo requiere, se le pueden leer los ítems para facilitar su comprensión.

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

Existen adaptaciones españolas de otros inventarios de Beck específicamente diseñados para adultos (p. ej., BDI-II; Beck, Steer y Brown, 2011), con procesos de desarrollo, validación y tipificación totalmente independientes del BYI-2 para niños y adolescentes.

El BYI-2 existe en versión impresa y versión informatizada.

1.19. Procedimiento de corrección:

La corrección es automatizada, mediante una plataforma de corrección on-line.

1.20. Puntuaciones:

Cuando la aplicación se realiza en papel, se deben introducir las respuestas en la plataforma on-line para realizar la corrección informatizada y generar el informe. Cuando la aplicación es informatizada, las respuestas se almacenan automáticamente.

Para cada inventario, se proporcionan puntuaciones directas (rango 0-60). La puntuacion directa total de cada uno de los cinco inventarios se transforma en una puntuación T (M= 50; DT=10). En cuatro de ellos, cuanto mayor es la puntuación T, mayor es el malestar psicológico que la persona refiere. La excepción es el inventario de autoconcepto (BSCI-Y), en el que cuanto menor es la puntuación, menor es el autoconcepto.

Se ofrecen, además, porcentajes acumulados de las puntuaciones T, que indican cuántas personas de la muestra de tipificación han obtenido la misma puntuación o una inferior. Las puntuaciones T obtenidas se asocian con niveles de gravedad con los que se puede clasificar a las personas evaluadas en función de rangos clínicos. Así, las puntuaciones T que se sitúan a varias desviaciones típicas por encima o por debajo de la media se asocian con niveles de significación clínica en la medida en que son más altas o más bajas que las puntuaciones del grupo normativo: en la media (T< 55), un poco alto (T= 55-59), bastante alto (T= 60-69) y muy alto (T ≥ 70).

No es posible corregir un inventario cuando hay más de dos ítems sin responder.

La corrección no se puede realizar manualmente. Las respuestas dadas por los respondientes deben introducirse en la plataforma de corrección on-line de Pearson Clinical & Talent Assessment y el sistema envía un perfil con los resultados obtenidos.

1.21. Escalas utilizadas:

Puntuaciones T (media 50 y desviación típica 10) y porcentajes acumulados.

1.22. Posibilidad de obtener informes automatizados:

Sí. La interpretación es clara, sencilla y visual, aunque requiere la interpretación de un experto. El informe incluye un perfil en el que se ilustran, con un gráfico de barras, las puntuaciones T en los cinco inventarios. Debajo de este, se incluye una tabla con la puntuación directa, la puntuación T, el porcentaje acumulado y el nivel de gravedad o rango clínico para cada inventario. Los informes escalan a los participantes en diferentes niveles, lo que facilita la comprensión y el análisis de las respuestas. Además, el manual recoge cuatro estudios de casos para ilustrar la interpretación de la evaluación (dos casos de niños en la franja de edad más baja y dos casos de adolescentes en la franja de edad más alta).

1.23. Tiempo estimado para la aplicación del test:

25-50 minutos en total (5 -10 minutos por inventario), tanto en aplicación individual como colectiva.

1.24. Documentación aportada por el editor:

Manual y cuadernillos de respuesta.

1.25. Precio de un juego completo de la prueba:

225 euros (manual, 25 cuadernillos de respuesta, 125 usos de aplicación y corrección on-line).

Puede adquirirse el manual de forma independiente por 60 euros.

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

39,95 euros (25 hojas de respuesta y 25 usos online en la plataforma Q-global) cada uno de los cinco inventarios que componen el BYI-2.

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

39,95 euros (25 hojas de respuesta y 25 usos online en la plataforma Q-global) cada uno de los cinco inventarios que componen el BYI-2.

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

175 euros: 25 cuadernillos de respuesta, 125 usos online en la plataforma Q-global (25 por inventario).

250 euros: uso ilimitado de aplicación y corrección del BYI-2 durante un año.

500 euros: uso ilimitado de aplicación y corrección del BYI-2 durante tres años.

750 euros: uso ilimitado de aplicación y corrección del BYI-2 durante cinco años.

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Buena	4
2.2. Calidad de la documentación aportada	Buena	4
2.3. Fundamentación teórica	Adecuada-Buena	3,5
2.4. Adaptación del test	Buena	4
2.5. Desarrollo de los ítems del test	Adecuada	3
2.6. Calidad de las instrucciones para el participante	Buena-Excelente	4,5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Buena-Excelente	4,5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Adecuada	3
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	Adecuada	3
2.11.1.2. Consultas a expertos	Se ha consultado a un pequeño número de expertos mediante un procedimiento sistematizado	3

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 tests, diferencias entre grupos, diseños experimentales o cuasi experimentales	
2.11.2.1.2. Tamaño de las muestras	Varios estudios con alguna muestra grande y otras pequeñas	4
2.11.2.1.3. Procedimiento de selección de las muestras	No se aporta información sobre las muestras utilizadas para estudiar la relación con otras variables Incidental en el caso de las muestras utilizadas para estudiar la relación con un criterio externo	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	Adecuada	3
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	Buena	4
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	Buena	4
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	No se aporta	---
2.11.2.1.8. Resultados de las diferencias intergrupo	Excelente	5

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	Muestra clínica española de niños y adolescentes que cumplían los criterios diagnósticos para trastornos depresivos, ansiedad, disruptivos, control de impulsos, de la conducta, de la conducta alimentaria y la ingesta de alimentos	
2.11.2.2.2. Calidad de los criterios empleados	Buena	4
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 pequeñas	2
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental	
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	Adecuada, pero con algunas carencias	2
2.11.3.2. Funcionamiento diferencial de los ítems	No se aporta	---

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:

La muestra de tipificación está compuesta por 703 personas con edades comprendidas entre 7:0 y 18:11 años, repartidos en tres grupos de edad (7:00-10:11=212; 11:0-14:11=245; 15:0-18:11=246), equilibrados por sexo y procedentes de distintas comunidades autónomas. Los porcentajes por edad y sexo se ajustan de forma satisfactoria a los porcentajes teóricos basados en el censo de la población española (INE, 2011). En términos generales, la muestra es valorada positivamente, mejora el muestreo del estudio original en Estados Unidos e incluye una interesante muestra clínica. En este sentido, sería deseable incluir información más detallada en el manual acerca del proceso de selección de los casos y las características de la muestra.

Las evidencias de validez basadas en el contenido se incluyen en el capítulo dedicado a explicar el desarrollo de la versión estadounidense (en la que se partía de 25 ítems por inventario para llegar a una versión final de 20) y de la adaptación española. En esta última, se contó con la colaboración de una traductora (que realizó la traducción del inglés al español) y de cinco expertos en psicología clínica infanto-juvenil (quienes analizaron la traducción y consensuaron la idoneidad de los ítems, las instrucciones y el formato de respuesta). Sería recomendable en futuras ediciones del manual detallar el proceso seguido para la creación de nuevos ítems, proporcionar coeficientes de concordancia entre los expertos, y especificar los ítems que se mantienen de la versión original, los de nueva creación y los de doble formulación.

Las evidencias de validez del BYI-2 que se proporcionan de forma específica en el manual son las basadas en la estructura interna de los inventarios, en las relaciones con otras medidas y en estudios realizados con muestra clínicas.

En cuanto a las evidencias basadas en la estructura interna, se examinaron las relaciones entre los ítems de los inventarios del test mediante estudios de consistencia interna, intercorrelación y de componentes principales. Por un lado, las correlaciones van en el sentido esperado: positivas entre los inventarios de depresión, ansiedad, ira y conductas disruptivas, y negativas entre estos cuatro inventarios y el autoconcepto. Por otro lado, se describen los resultados del análisis de componentes principales, con rotación oblicua (oblimin), en la muestra de la población general (para explorar el efecto de la edad, se analizan por separado la muestra de niños y la de adolescentes). Sobre estos análisis, sería recomendable incluir resultados detallados (p. ej., las saturaciones de los ítems en los factores de las distintas soluciones factoriales, valores propios, gráfico de sedimentación), asi como una discusión acerca del porcentaje de varianza explicado por la solución de cinco factores (un 33% para la muestra de niños y 36% para la de adolescentes, cuando lo esperable es que rondaran el 50%). De hecho, en futuras ediciones del test sería recomendable optar por otros análisis como análisis factorial confirmatorio o modelos exploratorios de ecuaciones estructurales.

Con respecto a las evidencias basadas en las relaciones con otras medidas, se presentan correlaciones entre las puntuaciones del tests con otros instrumentos que miden constructos relacionados estrechamente (validez convergente) y vagamente (validez discriminante). Más concretamente, se estudian las relaciones con otros instrumentos que miden depresión en la infancia (BDI-II, CDI), ansiedad (BAI, CMAS-R), conductas problemáticas (CASS:S), acoso escolar (RBVSS) y autoconcepto (PHCSCS, PHCSCS-2). Aunque los estudios de las relaciones con otras medidas se consideran uno de los puntos fuertes de este test, lo ideal en futuras ediciones seria realizar tales estudios con muestra española y proporcionar información detallada acerca del procedimiento de selección de todas las muestras. Finalmente, se consideran otra fortaleza de la prueba las evidencias basadas en muestras clínicas con datos españoles, al proporcionar información acerca de la capacidad discriminativa de los inventarios.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Coeficientes de fiabilidad y 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	Varios estudios con alguna muestra grande y otras pequeñas	4
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa	
2.12.3.3. Promedio de coeficientes de consistencia	Excelente	5

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Varios estudios con alguna muestra grande y otras pequeñas	4
2.12.4.2. Coeficientes de estabilidad	Excelente	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:

Se proporcionan tres indicadores de la fiabilidad de los inventarios BYI-2: (1) consistencia interna (i.e., alpha de Cronbach), (2) errores típicos de medida y (3) estabilidad temporal (i.e., fiabilidad test-retest). El valor promedio de los distintos coeficientes analizados se considera excelente.

En cuanto a la fiabilidad sería recomendable: (1) tener en cuenta en la redacción del manual que la fiabilidad no es del test, sino de las puntaciones (i.e., depende de la muestra analizada) y (2) añadir información adicional sobre otros índicadores de fiabilidad que soslayen las limitaciones del alfa Cronbach en respuestas ordinales, como por ejemplo alfa ordinal (Zumbo et al., 2007; 2012) u Omega.

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	Amplio rango de baremos en función de características relevantes	5
2.13.1.2. Tamaño de las muestras	Grande	4
2.13.1.3. Aplicación de estrategia de tipificación continua	Sí	
2.12.1.4. Procedimiento de selección de las muestras	Aleatorio, con muestras representativas de la población objetivo	
2.12.1.5. Actualización de baremos	Excelente	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:

El manual recoge, en primer lugar, algunos aspectos relacionados con la validez de la aplicación, haciendo hincapié en la necesidad de que los evaluadores se cercioren de las capacidades necesarias para que niños y adolescentes cumplimenten los inventarios por sí solos, así como para solventar posibles dificultades como la falta de sinceridad o de motivación para realizar la evaluación.

Los baremos del BYI-2 son uno de los puntos fuertes, basados en una muestra aleatoria de niños y adolescentes de población general, estratificada por sexo y edad.

La interpretación de las puntuaciones del test resulta muy sencilla. La corrección on-line proporciona para cada inventario las puntuaciones directas, las puntuaciones T (media 50 y desviación típica 10), el porcentaje acumulado e información sobre el 'rango clínico' o nivel de gravedad. Así, cuanto mayor es la puntuación T en el continuo, mayor es el malestar psicológico que el niño o el adolescente refiere. La excepción es el inventario del autoconcepto (BSCI-Y), en el que cuanto mayor es la puntuación, más positivo es el autoconcepto. Las puntuaciones T que se sitúan a varias desviaciones típicas por encima o por debajo de la media se asocian con determinados niveles de significación clínica en la medida en que son más altas o más bajas que las puntuaciones del grupo normativo. De este modo, se utilizan los siguientes rangos: en la media ($T < 55$), un poco alto ($T= 55-59$), bastante alto ($T= 60-69$) y muy alto ($T \geq 70$). Para ilustrar el proceso de interpretación, se exponen cuatro estudios de casos que resultan de gran ayuda. Sería interesante en futuras versiones comentar con detalle cómo se establecen los diferentes rango