

Evaluación de

Test de Habilidad en Inteligencia Emocional

THINEME

Prueba valorada en la ronda de evaluaciones del año 2023

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Test de Habilidad en Inteligencia Emocional (THINEME)
Autor	Isabel María Merchán, Juan de Dios González y José Luis Ramos
Autor de la adaptación española	N/A
Variable(s)	Inteligencia emocional
Áreas de aplicación	Psicología educativa. Psicología clínica.
Soporte	Papel y lápiz e Informatizado

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Adecuada	3
Fundamentación teórica	Buena	4
Adaptación	N/A (no aplicable)	--
Análisis de ítems	Adecuada	3
Validez: contenido	Buena	4
Validez: relación con otras variables	Adecuada, pero con algunas carencias	2.8
Validez: estructura interna	Buena	4
Validez: análisis del DIF	N/A (no aplicable)	-
Fiabilidad: equivalencia	N/A (no aplicable)	-
Fiabilidad: consistencia interna	Adecuada, pero con carencias	2.8
Fiabilidad: estabilidad	N/A (no aplicable)	-
Fiabilidad: TRI	N/A (no aplicable)	-
Fiabilidad inter-jueces	Adecuada, pero con carencias	2.5
Baremos e interpretación de puntuaciones	Buena	3

Comentarios generales

THINEME se define como una prueba de ejecución para evaluar inteligencia emocional (IE) en niños y niñas de 8 a 12 años para la medida de las cuatro ramas o dominios de habilidad emocional (percepción emocional, facilitación, comprensión y regulación) que plantea el modelo de Mayer y Salovey (1997). La prueba permite obtener tanto una puntuación global de IE como puntuaciones en las cuatro escalas que componen el instrumento:

- Percepción emocional: capacidad para identificar las emociones en uno mismo y en los demás, así como en los objetivos, el arte, la música y otros estímulos, de forma correcta y en el momento adecuado. La capacidad de expresarlos de forma correcta y en el momento adecuado y la facultad para reconocer la sinceridad de las emociones expresadas por los demás.
- Facilitación emocional: capacidad de generar, utilizar y sentir las emociones de la forma necesaria para comunicar los sentimientos o utilizarlos en otros procesos cognitivos
- Comprensión emocional: capacidad para comprender la información emocional, etiquetar las emociones, entender cómo se combinan y cómo evolucionan a lo largo de las transiciones interpersonales y apreciar los significados emocionales.
- Regulación emocional: Capacidad de abrirse a los sentimientos y de modularlos en uno mismo y en los demás, así como promover la comprensión personal y el crecimiento.

Su novedad radica en que son muy pocos los instrumentos para la medida de la inteligencia emocional en el campo educativo y más aún en pruebas de ejecución para el rango de edad para el que ha sido diseñado. Por tanto, la principal fortaleza de este instrumento es que proporciona una evaluación de una variable tan relevante como la inteligencia emocional, en unas edades muy tempranas. Esta fortaleza probablemente también sea el germen de las debilidades o aspectos a mejorar del instrumento relacionadas con sus propiedades psicométricas.

En relación con las evidencias de validación de contenido, el proceso inicial de depuración de ítems es correcto, con un estudio de 15 jueces que valoraron la pertinencia, la redacción y la suficiencia de los ítems a través de una escala tipo Likert de 5 puntos. Sin embargo, no está bien descrito el proceso de depuración de la prueba hasta llegar a la versión final. Se realizó un estudio piloto con un proceso de jueces en la primera versión de la prueba de 75 ítems, con un 90% de acuerdo interjueces. Fruto de esta revisión se redujo la escala a 69 ítems y después a 64, en una versión que incluye cambios sustantivos como la inclusión de imágenes. La adecuación de las imágenes se valoró mediante acuerdo interjueces, pero se encontraron valores bajos de acuerdo mediante kappa y W de Kendall. Falta información en el manual del test sobre la reducción final de la muestra a 30 ítems, aunque se apunta a motivos psicométricos. Parece ser que la muestra del estudio principal cumplimentó la versión de 64 ítems, aunque luego todos los resultados que se aportan se realizan sobre los 30 ítems finales. Es importante señalar que, aunque los ítems sufrieron modificaciones en su redacción por recomendación de los jueces, todavía hay algunos ítems cuya dificultad hace pensar que pueden estar midiendo otras competencias además de inteligencia emocional, lo que constituye un

problema de varianza irrelevante de constructo. En este sentido, es importante abordar o al menos tener en cuenta dos elementos significativos de la prueba: la comprensión lectora y la capacidad para interpretar las escenas visuales. La primera se relaciona con la competencia verbal, mientras que la segunda se refiere a la capacidad visual. Es esperable que los sujetos con deficiencias en cualquiera de estas competencias obtengan puntuaciones particularmente bajas en el instrumento. Por tanto, sería recomendable complementar esta prueba con otras que evalúen específicamente estas competencias verbales y visuales, para evitar interpretar como inteligencia emocional baja a niños con una buena inteligencia emocional alta, pero que presentan baja capacidad en tareas de procesamiento cognitivo.

En relación a las evidencias de validez basadas en la relación del test con otras variables, no es posible realizar estudios de validez convergente propiamente dichos, dado que no existen otros tests disponibles que evalúen la inteligencia emocional en niños de estas edades. Ante esta limitación, los autores han tomado la decisión de emplear a otro informante, en este caso, el profesor, para recabar información adicional. Sin embargo, se ha observado que las correlaciones obtenidas son bajas. Además, se dispone de una medida del rendimiento académico de los estudiantes (una pregunta realizada a los profesores), aunque no se ofrece detalle alguno sobre esta medida en particular en el manual. Por lo tanto, se sugiere que se proporcione más información al respecto para evaluar adecuadamente su idoneidad como contribución a la evidencia de validez referida al criterio.

En cuanto a las evidencias de validez interna, El AFC valida la estructura tetrafactorial del AFE, y presenta un buen ajuste, si bien los pesos de algunos ítems son algo bajos. Sería recomendable realizar un análisis de invarianza de medida para asegurar que las diferencias encontradas en sexo y edad se deben al rasgo evaluado y no a sesgos vinculados al instrumento.

Respecto a la fiabilidad, la baja consistencia interna encontrada en las cuatro subescalas, hace especialmente recomendable complementar la información sobre fiabilidad con otros procedimientos, como la fiabilidad test-retest o la función de información desde la Teoría de Respuesta al Ítem.

En conclusión, el objetivo de este instrumento es la evaluación de un constructo muy complejo, como es la Inteligencia Emocional en su concepción de habilidad, y en el que apenas hay referentes para las edades en la que se pretende medir. Esta gran dificultad explica en parte que algunas de las propiedades psicométricas estén por debajo de lo recomendado, tanto en la discriminación o pesos factoriales, como en la fiabilidad de las escalas que lo componen. A pesar de estas dificultades, el test proporciona información que puede resultar útil para orientadores y educadores, tal y como se muestra de forma muy clara en los casos ilustrativos del manual.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Test de Habilidad en Inteligencia Emocional (THINEME)

1.2. Nombre del test en su versión original:

N/A

1.3. Autor/es del test original:

Isabel María Merchán, Juan de Dios González y José Luis Ramos

1.4. Autor/es de la adaptación española:

N/A

1.5. Editor del test en su versión original:

GIUNTI Psychometrics

1.6. Editor de la adaptación española:

N/A

1.7. Fecha de publicación del test original:

2023

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

N/A

1.9. Fecha de la última revisión del test:

2023

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

Inteligencia emocional.

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

THINEME se define como una prueba de ejecución para evaluar inteligencia emocional (IE) en niños y niñas de 8 a 12 años para la medida de las cuatro ramas o dominios de habilidad emocional (percepción emocional, facilitación, comprensión y regulación) que plantea el modelo de Mayer y Salovey (1997). Su novedad radica en que son muy pocos los instrumentos para la medida de la IE en el campo educativo y más aún en pruebas de ejecución para el rango de edad para el que ha sido diseñado.

La prueba permite obtener tanto una puntuación global de IE como puntuaciones en las cuatro escalas que componen el instrumento y que se definen a continuación:

- Percepción emocional: capacidad para identificar las emociones en uno mismo y en los demás, así como en los objetivos, el arte, la música y otros estímulos, de forma correcta y en el momento adecuado. La capacidad de expresarlos de forma correcta y en el momento adecuado y la facultad para reconocer la sinceridad de las emociones expresadas por los demás.
- Facilitación emocional: capacidad de generar, utilizar y sentir las emociones de la forma necesaria para comunicar los sentimientos o utilizarlos en otros procesos cognitivos
- Comprensión emocional: capacidad para comprender la información emocional, etiquetar las emociones, entender cómo se combinan y cómo evolucionan a lo largo de las transiciones interpersonales y apreciar los significados emocionales.
- Regulación emocional: capacidad de abrirse a los sentimientos y de modularlos en uno mismo y en los demás, así como promover la comprensión personal y el crecimiento.

En los casos de que se detecten bajas puntuaciones de IE, la información obtenida puede ser utilizada para diseñar e implementar intervenciones para el desarrollo emocional. Las posibilidades de utilización se expanden desde la orientación educativa, al ámbito clínico, psicopedagógico, detección de perfiles emocionales en niños y niñas de manera individual o colectiva., así como con necesidades específicas derivados de distintos supuestos –trastorno por déficit de atención e hiperactividad, alta capacidad intelectual o discapacidad cognitiva, trastornos del espectro autista, trastornos de aprendizajes, trastornos de la lectoescritura, trastornos graves de conducta, problemas de salud mental-, etc.

1.12. Áreas de aplicación:

Psicología educativa

Psicología clínica

1.13. Formato de los ítems:

Elección múltiple

1.14. Número de ítems:

Escala de Regulación: 8 ítems

Escala de Facilitación: 7 ítems

Escala de Percepción: 8 ítems

Escala de Comprensión: 7 ítems

Total: 30 ítems

1.15. Soporte:

Papel y lápiz e Informatizado.

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

Nivel B

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

Thineme es un test dirigido a niños y niñas con edades entre los 8 y 12 años. El nivel educativo correspondiente a esas edades es segundo y tercer ciclo de Educación Primaria.

El test es aplicable en niños con desarrollo típico, y los autores informan de su posible aplicación para niños con trastorno del desarrollo, indicando como posible acomodación la presentación oral por parte del evaluador en caso de dificultades de comprensión lectora.

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

El test tiene una versión en lápiz y papel y otra online. Únicamente se diferencian en la forma de aplicación.

1.19. Procedimiento de corrección:

Automatizado por ordenador.

1.20. Puntuaciones:

Puntuaciones directas: se calculan sumando los valores numéricos de las respuestas dadas a cada ítem. Puesto que hay cinco opciones de respuesta, se asigna a dos de ellas cero puntos, mientras que las otras tres son valoradas con 1, 2 y 3 puntos.

No existe ningún procedimiento de control de respuesta al azar y el sistema de puntuaciones se aplica tanto para las subescalas como para la puntuación global.

Para la interpretación de las puntuaciones se utilizan centiles y puntuaciones T.

El criterio de las asignaciones de puntuación a cada una de las alternativas de respuesta de los ítems no se hace explícito en el manual. Exige corrección con la plataforma online para grabar las respuestas, de acuerdo a unos códigos, proporcionando ya la puntuación final en forma de puntuaciones T y centiles.

1.21. Escalas utilizadas:

Percentiles y puntuaciones T (media 50 y desviación típica 10).

1.22. Posibilidad de obtener informes automatizados:

En el manual se indica que con la corrección informatizada se genera un perfil en el que se muestra la posición del individuo en relación con el grupo normativo, y además se genera un informe con una clave para su interpretación, y en algunos casos orientaciones para una intervención específica.

1.23. Tiempo estimado para la aplicación del test:

Tanto en la aplicación individual como en la colectiva el tiempo de aplicación varía de 30 a 60 minutos en función de la edad y la agilidad del estudiante evaluado.

1.24. Documentación aportada por el editor:

Manual y cuadernillo de respuestas.

1.25. Precio de un juego completo de la prueba:

69,90 €. Incluye manual técnico, hoja de respuestas (10 uds.) y usos de corrección (10 usos). Fecha de consulta de precios: 11-01-2024.

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

52,90 € (10 cuadernillos de respuesta + uso de corrección). Fecha de consulta de precios: 11-01-2024.

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

El cuadernillo incluye hoja de respuestas.

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

En la administración por lápiz y papel, junto al Cuadernillo se incluyen los usos de corrección. En la administración online se ofrece por 52,00 € 10 usos online, que permiten auto aplicación y obtener un informe interpretativo de la prueba, si bien requiere adquirir el manual técnico y cuaderno de láminas. Fecha de consulta de precios: 11-01-2024.

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	Adecuada	3
2.3. Fundamentación teórica	Buena	4
2.4. Adaptación del test	N/A	-
2.5. Desarrollo de los ítems del test	Adecuada	3
2.6. Calidad de las instrucciones para el participante	Adecuada	3
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Adecuada	3
2.8. Facilidad para registrar las respuestas	Buena	4
2.9. Bibliografía del manual	Adecuada - Buena	3.5
2.10. Datos sobre el análisis de los ítems	Adecuados	3

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	Buena. Consultados 15 expertos.	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 Correlaciones con evaluaciones específicas realizadas por el profesorado de los participantes y correlaciones con el rendimiento académico.	
2.11.2.1.2. Tamaño de las muestras	Un estudio con una muestra grande	3
2.11.2.1.3. Procedimiento de selección de las muestras	Los autores informan de que se realiza un muestreo por conglomerados en centros seleccionados aleatoriamente de la Comunidad Autónoma de Extremadura. La muestra incluye centros públicos y concertados de zonas céntricas y periféricas de ciudades y de zonas rurales, acogiendo así alumnos de estratos socioeconómicos y culturales diversos. Falta información sobre las proporciones de los diversos estratos para valorar la representación de lo que parece ser un muestreo estratificado en combinación con el muestreo por conglomerados.	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	Se utilizan dos preguntas elaboradas ad hoc que responden los profesores del alumnado. No hay evaluación de la calidad de esta medida.	-
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	Adecuada, pero con algunas carencias	2.5
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	N/A	-

2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	N/A	--
2.11.2.1.8. Resultados de las diferencias intergrupo	Adecuados	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	Los autores no presentan evidencias de validez de criterio como tal. Sin embargo, en el apartado de evidencias de validez basadas en el contenido, presentan resultados de la correlación entre la puntuación del test total y de las subescalas con una medición del rendimiento académico, que consiste en una pregunta realizada al profesor, con cuatro opciones ordenadas de respuesta.	
2.11.2.2.2. Calidad de los criterios empleados	No se aporta información sobre la calidad de la medida del rendimiento académico utilizada.	-
2.11.2.2.3. Relación temporal entre test y criterio	Concurrente.	
2.11.2.2.4. Tamaño de las muestras	Un estudio con una muestra grande.	3
2.11.2.2.5. Procedimiento de selección de las muestras	Los datos se basan en la muestra general aplicada, ya descrita en otros apartados.	
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	Buena	4
2.11.3.2. Funcionamiento diferencial de los ítems	N/A	-

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	Sí. En caso de dificultades en comprensión lectora se recomienda la administración oral en	

2.11.5. Comentarios generales sobre evidencias de validez:

Se valora cada una de las evidencias presentadas.

a) En relación al contenido.

En el proceso de creación de la prueba se realiza un estudio piloto con varias fases:

- La valoración de 15 jueces expertos que evaluaron la construcción de los 75 ítems de la primera versión, teniendo en cuenta la pertinencia, la redacción y la suficiencia, a través de una escala de 5 puntos. Respecto a la fiabilidad Inter jueces, se encuentran resultados de acuerdo por encima del 90% (no se aportan índices).
- La aplicación del instrumento (ahora de 69 ítems) a una muestra de 200 niños, que lleva a una depuración de algunos ítems, quedando una versión de 64 ítems.

Si bien el estudio piloto sigue las recomendaciones para intentar muestrear el contenido, la versión final del test queda reducida a 30 ítems, por lo que sería deseable ampliar la información en el manual sobre este proceso de depuración basado en motivos psicométricos.

Una de las recomendaciones de los expertos fue incluir imágenes para ilustrar las situaciones de los ítems a modo de comic. Estas imágenes fueron evaluadas por 8 jueces que tenían que realizar dos tareas. En la primera, se mostraron las imágenes que acompañan a las situaciones junto con los enunciados correspondientes y se les preguntó si la imagen es coherente con la situación presentada (escala de respuesta: 1. Muy incoherente; 2. Poco coherente; 3. Bastante coherente; y 4. Perfectamente coherente). La fiabilidad interjueces de los juicios de valoración de adecuación de las imágenes con la situación presentada, muestra valores de acuerdo bajos (Kappa = 0.189 y W de Kendall= 0.31). En la segunda tarea se valoraron 4 de los 5 ítems de la Situación 4, preguntando a los jueces si "la imagen representa la emoción de forma clara", con una escala de valoración de 1 a 4. En este caso, los valores de acuerdo fueron altos (Kappa = 0.702 y W de Kendall= 0.895).

b) Evidencias de validez basadas en la relación con otras variables. En el manual, las dos primeras se presentan inapropiadamente como evidencias de validez de contenido.

- a. Se utiliza como evidencia de validez convergente dos preguntas elaboradas ad hoc que responden los profesores del alumnado. No hay valoración de la calidad de esta medida y las correlaciones encontradas con la prueba son bajas (todas inferiores a .25).
- b. Se informa de las correlaciones del test (total y subescalas) con una medida del rendimiento académico, lo que se podría considerar una evidencia de validez de criterio. Convendría incluir información sobre la medida de rendimiento que se utiliza en el manual del test. Se trata de una pregunta con cuatro opciones de respuesta: 1) No alcanza los objetivos, 2) alcanza los objetivos con dificultad o ayuda, 3) Alcanza los objetivos esperados y 4) supera ampliamente los objetivos.
- c. Análisis de las diferencias en función de variables sociodemográficas (sexo y edad). Las diferencias por sexo se valoraron mediante contrastes de medias (t medidas independientes), y medidas del tamaño del efecto (d de Cohen), encontrando diferencias significativas en dos de las dimensiones (facilitación y

percepción), si bien el tamaño del efecto es pequeño. Para valorar las diferencias por edad, se utilizó un análisis de varianza (ANOVA) y pruebas de eta cuadrado, encontrando que hay diferencias en todos los grupos de edad excepto en los dos más mayores (11 y 12 años).

- c) Evidencias de validez basadas en la estructura interna del test. Se ha puesto a prueba mediante análisis factorial exploratorio (AFE) y análisis factorial confirmatorio (AFC) la estructura teórica de cuatro factores. El procedimiento de dividir la muestra en dos mitades aleatorias para cada análisis es adecuado. El AFC valida la estructura tetrafactorial del AFE, y presenta un buen ajuste, si bien la saturación de algunos ítems es algo baja. No se realizan pruebas de DIF. Sería necesario realizar un análisis de invarianza de medida para asegurar que las diferencias encontradas en sexo y edad se deben al rasgo evaluado y no a sesgos vinculados al instrumento.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Un único coeficiente de fiabilidad aplicado a cada subescala y a la escala total.	

2.12.2. Equivalencia formas paralelas:

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

2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Adecuada. Un estudio con una muestra grande (1366 estudiantes)	3
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa	
2.12.3.3. Promedio de coeficientes de consistencia	Adecuado, pero con algunas carencias	2.5

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	N/A	-
2.12.4.2. Coeficientes de estabilidad	N/A	-

2.12.5. Cuantificación de la precisión mediante TRI:

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

2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	Coeficiente kappa y coeficiente de concordancia W de Kendall	
2.12.6.2. Promedio de los coeficientes	Adecuado, pero con algunas carencias	2

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

Las evidencias de fiabilidad de las puntuaciones en la prueba sometida a valoración se han recogido sobre una única muestra de tamaño amplio (1366 escolares), que se corresponde con la muestra de baremación.

La fiabilidad como consistencia interna se estima con un único índice, el coeficiente alfa. El valor de este coeficiente para el total de la escala es bueno (0.83) pero, por subescalas, los valores de consistencia interna son demasiado bajos (entre 0.57 y 0.64). No se proporciona el valor del coeficiente alfa por sexo o por edad.

En cuanto al análisis de ítems se presenta la correlación ítem-test corregida y el valor del coeficiente alfa si se elimina el elemento. En cuanto a los valores de discriminación de los ítems, hay 9 ítems (de los 30 que conforman la escala) con un valor de correlación ítem-test corregida por debajo de 0.30 (dos en la dimensión de regulación, uno en la dimensión de facilitación, tres en la dimensión de percepción y tres en la dimensión de comprensión). La baja discriminación de los ítems y el reducido número de ítems por dimensión probablemente sean los responsables de los bajos valores de consistencia interna observados en las cuatro dimensiones (subescalas) analizadas.

No valoran la fiabilidad en cuanto a la estabilidad de las puntuaciones, lo cual podría ser una información muy relevante dado que los valores de consistencia interna no son altos.

No se aporta información sobre los resultados TRI, ya que no encuentran buen ajuste al modelo empleado (no se informa de cuál es).

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	Un baremo aplicable a la población objetivo con cierta precaución (no se separan por sexo)	2.5
2.13.1.2. Tamaño de las muestras	Suficiente - Moderado (por grupo de edad, entre 230 y 356)	2.5
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	Los autores informan de que se realiza un muestreo por conglomerados en centros seleccionados aleatoriamente de la Comunidad Autónoma de Extremadura. La muestra abarca tanto centros públicos como concertados ubicados en áreas tanto urbanas como rurales, incluyendo alumnos pertenecientes a diversos estratos socioeconómicos y culturales. Falta información sobre el número de centros participantes y sobre las proporciones de los diversos estratos, para valorar la representación de lo que parece ser un muestreo estratificado en combinación con el muestreo por conglomerados.	
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	N/A	-
2.13.2.2. Procedimiento empleado para fijar los puntos de corte	N/A	
2.13.2.3. Procedimiento de obtención del acuerdo inter-jueces	N/A	
2.13.2.4. Valor del coeficiente de acuerdo inter-jueces	N/A	--

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

Los autores informan de que se realiza un muestreo por conglomerados en centros seleccionados al azar de la Comunidad Autónoma de Extremadura. Los autores consideran que “la muestra podría ser representativa de la población escolar española puesto que se incluyeron en él centros públicos y concertados de zonas céntricas y periféricas de ciudades y de zonas rurales, acogiendo así alumnado de estratos socioeconómicos y culturales diversos de manera proporcionada”. Falta información sobre las proporciones de los diversos estratos para valorar la representación de lo que parece ser un muestreo estratificado en combinación con el muestreo por conglomerados. En cualquier caso, toda la muestra proviene exclusivamente de Extremadura, lo que cuestiona la afirmación previa sobre su representatividad para toda la población escolar española.

La muestra de tipificación está formada por 1366 participantes, aproximadamente la mitad niños y la mitad niñas. No se proporciona información sobre si esta proporción de género se mantiene en los diferentes grupos de edad de 8 a 12 años. Los tamaños muestrales en los cinco grupos de edad varían de 125 a 356 sujetos, si bien las normas se ofrecen para cuatro grupos de edad, al unirse en un grupo el alumnado de 11 y 12 años. Los análisis realizados muestran diferencias estadísticamente significativas por sexo en dos de las cuatro subescalas. Aunque el tamaño del efecto es pequeño, sería recomendable disponer de baremos por separado para niños y niñas.

Para la interpretación normativa de las puntuaciones se proporcionan percentiles y puntuaciones T, con una clasificación en función de estas últimas en: muy bajo (< 34), bajo (35-44), medio (45-55), alto (56-65) y muy alto (>65). Los autores consideran que su prueba es un test referido al criterio, pero no se aportan puntos de corte (más allá de la interpretación de puntuaciones T), ni estudios relativos al establecimiento de puntos de corte (juicio de expertos o investigaciones empíricas) que permitan realizar una interpretación criterial.