

Evaluación de

Batería de Evaluación del Bienestar Educativo

∞ Batería EBE ∞

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Batería de Evaluación del Bienestar Educativo (Batería EBE)
Autores	Jesús de la Fuente Arias y José Manuel Martínez-Vicente
Autor de la adaptación española	-
Variable(s)	<ul style="list-style-type: none">- Escala para la Evaluación de la Regulación Interna y externa del Aprendizaje. Compuesta por 2 subescalas: Subescala de Regulación Personal y Subescala de Regulación del Contexto (formadas a su vez por tres factores cada una de ellas)- Escala Abreviada para la Evaluación de la Regulación del Proceso Enseñanza-Aprendizaje. Compuesta por 2 subescalas: Enseñanza Reguladora y Aprendizaje Autoregulado. La primera se compone a su vez por 4 factores y la segunda por 6 factores- Escala de Salud Académica. Compuesta por 2 subescalas: Salud Física y Salud Psicológica <p>La combinación de las tres escalas permite la obtención de un Índice de Bienestar Educativo</p>
Áreas de aplicación	Psicología educativa
Soporte	Administración y corrección automática vía internet

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Buenos	4
Fundamentación teórica	Buena	4
Adaptación	N/A (no aplicable)	-
Análisis de ítems	Adecuado	3
Validez: contenido	Adecuada	3
Validez: relación con otras variables	Adecuada, pero con algunas carencias- Adecuada	2,5
Validez: estructura interna	Adecuada, pero con algunas carencias	2,5
Validez: análisis del DIF	Adecuada	3
Fiabilidad: equivalencia	N/A (no aplicable)	-
Fiabilidad: consistencia interna	Adecuada-Buena	3,5
Fiabilidad: estabilidad	N/A (no aplicable)	-
Fiabilidad: TRI	N/A (no aplicable)	-
Fiabilidad inter-jueces	N/A (no aplicable)	-
Baremos e interpretación de puntuaciones	Adecuados	3,33

Comentarios generales

La Batería de Evaluación del Bienestar Educativo (Batería EBE) es una herramienta compuesta por tres escalas desarrolladas anteriormente por los autores: 1) Escala para la Evaluación de la Regulación Interna y Externa del Aprendizaje (SRL-ERL) (de la Fuente, 2022) formada por 36 ítems que se agrupan en dos subescalas de 18 ítems cada una, subescala de Regulación Personal y subescala de Regulación del Contexto. Cada una de estas subescalas a su vez se configura según tres factores; 2) Escala Abreviada para la evaluación interactiva del proceso de Enseñanza-Aprendizaje (EIPEA-Abreviada) (de la Fuente y Martínez-Vicente, 2004). Está formada por dos subescalas de 27 ítems (Enseñanza Reguladora) y 32 ítems (Aprendizaje Autorregulado). En el caso de la primera subescala se configura en 4 factores y la segunda subescala se configura en 6 factores, y 3) Escala de Salud Académica compuesta por 2 subescalas: Salud Física y Salud Psicológica. Las tres escalas se combinan para obtener un índice general de Bienestar Educativo (IBE) que es una de las principales aportaciones de la Batería EBE.

En global, se trata de un autoinforme de 103 ítems con respuestas tipo Likert. La batería es de utilidad para detectar factores protectores y de riesgo relacionados con el proceso de enseñanza-aprendizaje y el bienestar escolar/académico en población escolar entre 17 a 21 años. Los materiales son adecuados, completos y fáciles de comprender, aunque se agradecería que estuvieran organizados de forma más lógica y sistemática. Las instrucciones para una adecuada administración de la batería son claras y sencillas. La corrección es automática dado que la aplicación es online.

En cuanto a las fuentes de validez, son adecuadas, pero con ciertas carencias. Se proporcionan evidencias de validez del contenido, de relación del test con una escala de bienestar psicológico, y de la estructura interna de las puntuaciones en las tres escalas que configuran la Batería EBE. Sin embargo, algunas decisiones sobre la estructura interna no están adecuadamente justificadas en el manual. Sería recomendable aportar información sobre el funcionamiento diferencial de los ítems respecto a alguna variable de interés, como podría ser el sexo de los participantes u otra variable relevante (información que fue compartida con el coordinador de la revisión pero no incluida en el manual). Dado que entre los objetivos de la batería se encuentra evaluar factores relativos a la experiencia de bienestar escolar/académico estos factores estarían relacionados con variables del contexto de aprendizaje (p.ej. tipo de centro, características curriculares, etc.), sería recomendable obtener evidencias de validez de relación de las puntuaciones de la batería con este tipo de variables. También sería de utilidad obtener evidencias de diferencias en función del sexo, la edad o nivel educativo. Por otro lado, sería recomendable obtener evidencias de la capacidad de las puntuaciones de la batería para el pronóstico del rendimiento académico o de alguna otra variable relevante en el contexto educativo.

La fiabilidad se analiza en el contexto de la Teoría Clásica de los Tests (TCT) y se ofrecen diversos indicadores de consistencia interna (como el alfa de Cronbach), a partir de los cuales se concluye que la fiabilidad de las puntuaciones es por lo menos adecuada y en algunos casos buena o excelente, dependiendo de las dimensiones evaluadas. Se recomienda aportar información de la estabilidad de las puntuaciones obteniendo la fiabilidad test-retest. Adicionalmente, para la escala de Salud se recomienda completar la información con otras estimaciones de fiabilidad que no penalicen tanto a escalas compuestas por un reducido número de ítems.

Los baremos son adecuados, pero con ciertas limitaciones. La muestra utilizada

para la baremación es de estudiantes de Bachillerato y universitarios, las edades están comprendidas entre 17 y 24 años, con una edad media de 22,9 años y una desviación típica de 6,2. El número total de participantes fue de 1.268 y la mayoría (68%) fueron mujeres. No se proporciona información de las fechas de aplicación de los instrumentos, aunque por la información que se aporta en el manual se puede deducir que se aplicaron con posterioridad al año 2017. No se incluye información sobre otras características sociodemográficas de los participantes en el estudio, ni de los centros o diversidad territorial. Se detecta una cierta incongruencia acerca del rango de edad para la aplicación de la batería (entre 17 a 21 años), la media de edad de los participantes sobre los que se obtuvieron los datos de baremación (22,9 años) y la construcción de dos baremos diferenciados en función de la edad (grupo normativo de 17 a 20 y grupo normativo de 21 a 24 años). Dado que no se aportan datos sobre la existencia o no de posibles diferencias en función del sexo, no es posible valorar si es necesario establecer baremos en función de esta variable. Sin embargo, sería recomendable profundizar en esta cuestión.

En líneas generales, la Batería EBE presenta varios puntos fuertes que contribuyen a su utilidad en la evaluación del bienestar educativo. Sus puntos fuertes son:

- El manual de la Batería EBE ofrece una descripción detallada y bien fundamentada de su marco teórico. La inclusión de esta información es fundamental, ya que proporciona a los usuarios un entendimiento sólido de las bases conceptuales del instrumento.
- Destacar la facilidad de aplicación, ya que la plataforma web para administrar el test es muy cómoda y fácil de entender.
- El informe de los resultados es claro y contiene un perfil con elementos gráficos e informes completos, lo que es esencial para que los usuarios y profesionales comprendan la interpretación de los resultados de la Batería EBE. Esta característica fortalece la utilidad práctica de la herramienta.
- La oferta de una herramienta de corrección informatizada asegura la precisión de los resultados, minimizando posibles errores humanos. De igual manera, la aplicación online facilita el uso de la Batería EBE en el contexto de aplicaciones colectivas.
- Se observan valores notables de fiabilidad, medida con Alfa de Cronbach, en dos de las tres Escalas que componen la Batería: Escala para la Evaluación de la Regulación Interna y Externa del Aprendizaje (SRL-ERL) y Escala Abreviada para la evaluación interactiva del proceso de Enseñanza-Aprendizaje (EIPEA-Abreviada).

Sin embargo, sus puntos débiles, carencias o áreas de mejora son las siguientes:

- La falta de información detallada en el manual sobre el desarrollo inicial de las escalas que componen la batería. Aunque esta pueda obedecer a una estrategia editorial, los usuarios de la prueba deberían disponer de información, aunque sea muy resumida, de este aspecto.
- No se aplica ningún control para evitar sesgos de respuesta, lo que podría estar distorsionando la estructura del cuestionario, así como sus resultados.

- El manual no informa sobre características sociodemográficas relevantes de la muestra normativa (i.e. nivel académico, tipo de centro educativo, etc.) y tampoco aporta detalle sobre cómo fueron recabados los participantes y si se realizó algún mecanismo de control de calidad sobre los datos recabados.
- En el manual, se hubiera agradecido en algunos puntos más detalle sobre la justificación de algunas decisiones, especialmente, con respecto a las evidencias de validez y al análisis de la fiabilidad.
- La fiabilidad de la escala de Salud está más comprometida por el reducido número de ítems. Otras estimaciones de fiabilidad podrían ser recomendables. La inclusión de indicadores basados en la TRI permitiría conocer la precisión en diferentes niveles del rasgo. También sería recomendable considerar estimaciones de la estabilidad temporal. Finalmente, y esto es muy relevante, en la versión del Manual evaluada en esta revisión no se ha aportado ninguna estimación de la fiabilidad para el Índice de Bienestar Subjetivo (IBE) que es una de las principales aportaciones de la batería EBE.
- Las evidencias de validez presentan distintas limitaciones. Por ejemplo, no se presenta información detallada de los datos recabados de los expertos para la validez de contenido, el análisis de DIF presente en el Manual actual de la prueba es inadecuado y en el Manual actual tampoco se consideran evidencias de validez discriminante, entre otras.
- Finalmente, resulta sorprendente que el ámbito de aplicación fuera tan reducido. Parece razonable pensar que la Batería EBE podría adaptarse a población adolescente sin muchas dificultades, al menos para cubrir el rango correspondiente al segundo ciclo de la Educación Secundaria Obligatoria (14-16 años).

En resumen, la Batería EBE ofrece una herramienta sólida para evaluar el Bienestar Educativo y sus diferentes componentes, aunque presenta algunas limitaciones que podrían ser abordadas en futuros estudios o ediciones del instrumento.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Batería de Evaluación del Bienestar Educativo (Batería EBE)

1.2. Nombre del test en su versión original:

N/A

1.3. Autor/es del test original:

Jesús de la Fuente Arias y José Manuel Martínez-Vicente

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:

2024

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

N/A

1.9. Fecha de la última revisión del test:

2024

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

Bienestar en contextos educativos, Proceso de Enseñanza-Aprendizaje

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

La Batería EBE es un instrumento derivado del "Índice de Experiencia de Bienestar Escolar y Académico" (de la Fuente y Martínez-Vicente, 2024) que permite evaluar múltiples dimensiones relacionadas con el bienestar en el contexto educativo. Este índice abarca factores de regulación personal y contextual, la calidad del proceso de enseñanza-aprendizaje y el nivel de salud académica física y psicológica, los cuales se consideran determinantes para el bienestar psicológico y el flourishing en el entorno académico.

Consta de tres instrumentos, lo que permite una evaluación global y de cada uno de sus factores:

1. Escala para la Evaluación de la Regulación Interna y Externa del Aprendizaje (SRL-ERL) (de la Fuente, 2022). Evalúa el nivel de regulación personal y contextual del alumnado en los procesos educativos. Está formada por 36 ítems que se agrupan en dos subescalas de 18 ítems cada una, Subescala de Regulación Personal y Subescala de Regulación del Contexto. Cada una de estas subescalas a su vez se configura según tres factores.
2. Escala Abreviada para la evaluación interactiva del proceso de Enseñanza-Aprendizaje (EIPEA-Abreviada) (de la Fuente y Martínez-Vicente, 2004). Evalúa la regulación del proceso de enseñanza y del aprendizaje desde la perspectiva del alumnado. Está formada por dos subescalas de 27 ítems (Enseñanza Reguladora) y 32 ítems (Aprendizaje Autorregulado). En el caso de la primera subescala se configura en 4 factores (Comportamiento general de enseñanza del profesorado, Enseñanza reguladora específica, Preparación para el aprendizaje y Evaluación reguladora). La segunda subescala se configura en 6 factores (Satisfacción con el proceso de aprendizaje, Estrategias de aprendizaje y autorregulación, Comportamiento autorregulado en el aula, Motivación hacia el aprendizaje, Autoevaluación y planificación del aprendizaje y Estrategias de aprendizaje).
3. Escala de Salud Académica. Evalúa conductas relacionadas con la salud física y psicológica. Está configurada por dos subescalas: Salud Física (5 ítems) y Salud Psicológica (3 ítems).

Cada una de las Escalas proporciona un índice: Índice de Regulación Personal-Contextual (IRPC), Índice Regulación Enseñanza-Aprendizaje (IREA) y el Índice de Salud Académica (ISA). Finalmente, la batería proporciona un índice de Bienestar Educativo (IBE), que representa el grado de protección o de riesgo de un sujeto o grupo en el contexto educativo.

1.12. Áreas de aplicación:

Psicología educativa

1.13. Formato de los ítems:

Respuesta graduada / tipo Likert

1.14. Número de ítems:

- Escala para la Evaluación de la Regulación Interna y Externa del Aprendizaje (SRL-ERL) (de la Fuente, 2022). Consta de 36 ítems.
- Escala Abreviada para la evaluación interactiva del proceso de Enseñanza-Aprendizaje (EIPEA-Abreviada) (de la Fuente y Martínez-Vicente, 2004). Consta de 59 ítems.
- Escala de Salud Académica. Consta de 8 ítems.

En total, 103 ítems.

1.15. Soporte:

Informatizado

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

Prueba de nivel B, según la página web de la editorial: pruebas que requieren formación y experiencia en el uso de test, conocimientos de psicometría y experiencia en el área de evaluación objeto del test

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

El ámbito de aplicación especificado es entre 17 y 21 años. Está orientado a personas en el contexto educativo. Aunque no se proporciona información sobre esta cuestión de forma explícita en el manual, es perfectamente aplicable a minorías étnicas, personas con discapacidad o grupos clínicos, siempre y cuando su edad esté en el rango de aplicación, formen parte de un contexto educativo y sean capaces de comprender enunciados simples.

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

La única versión del test es informatizada, tanto en su aplicación como en la corrección. La batería se responde a través de un navegador web, por lo que los únicos requisitos son tener un dispositivo con conexión a internet.

1.19. **Procedimiento de corrección:**

Automatizado por ordenador (existe software de corrección o plataformas de corrección on-line)

1.20. **Puntuaciones:**

No se expone con gran detalle el procedimiento exacto de obtención de las puntuaciones. Se menciona que: "El proceso de corrección se realiza de manera automática en la plataforma online [...] La Batería EBE proporciona una puntuación para cada subescala, así como las puntuaciones de los factores que las componen y los Índices que se derivan". El documento habla de "puntuaciones directas", que son transformadas a "puntuaciones normalizadas" y en los informes se presentan los tres tipos de puntuaciones. Sin embargo, y a juzgar por el material, el informe gráfico representa la puntuación del sujeto en percentiles, indicando valores "bajos" a valores inferiores al percentil 25, y valores "altos" puntuaciones superiores al percentil 75. No se especifican métodos para la corrección de la probabilidad de responder al azar, ni para prevenir aparición de posibles sesgos de respuesta.

1.21. **Escalas utilizadas:**

Puntuaciones basadas en percentiles y puntuaciones T

1.22. **Posibilidad de obtener informes automatizados:**

SI. Se ofrecen dos tipos de informes: gráfico y/o interpretativo.

El informe gráfico representa las puntuaciones percentiles en cada índice, escala y subescala, organizando los aspectos evaluados en filas. El informe es muy completo y vistoso, permitiendo de un vistazo localizar aspectos que destaquen de forma positiva o negativa.

En términos generales, es un informe de calidad, estructurado y claro.

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

60 minutos

1.24. **Documentación aportada por el editor:**

Manual y clave de acceso para la administración y corrección por internet

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

48 euros. Fecha de consulta: 25/01/2025

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

N/A

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

N/A

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

1 uso de autoaplicación on line 3,50 euros y 10 usos de autoaplicación on line 33,50 euros. Fecha de consulta 25/01/2025

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Buena-Excelente	4,5
2.2. Calidad de la documentación aportada	Adecuada-Buena	3,5
2.3. Fundamentación teórica	Buena	4
2.4. Adaptación del test*	N/A	N/A
2.5. Desarrollo de los ítems del test	Adecuado pero con limitaciones	2
2.6. Calidad de las instrucciones para el participante	Buena	4
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Buena	4
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Buena-Excelente	4,5
2.10. Datos sobre el análisis de los ítems	Adecuado	3

*Muestras empleadas en la adaptación:

	Muestras locales (seleccionadas de la población objetivo)	Muestras mixtas (algunas locales y otras internacionales)	Muestras internacionales (no proceden de la población objetivo)
Análisis de ítems	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validez: contenido	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validez: relación con otras variables	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validez: estructura interna	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Validez: análisis del DIF	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fiabilidad: equivalencia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fiabilidad: consistencia interna	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fiabilidad: estabilidad	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fiabilidad: TRI	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fiabilidad inter-jueces	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Baremos e interpretación de puntuaciones	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--	--------------------------	--------------------------	--------------------------

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	Adecuada	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 entre test	
2.11.2.1.2. Tamaño de las muestras	Un estudio con una muestra grande (N>500)	3
2.11.2.1.3. Procedimiento de selección de las muestras	Incidental	
2.11.2.1.4. Calidad de los test empleados como criterio o marcador	Adecuada, pero con algunas carencias- Adecuada	2,5
2.11.2.1.5. Promedio de las correlaciones con otros test que miden constructos similares	Adecuada, pero con algunas carencias	2
2.11.2.1.6. Promedio de las correlaciones con otros test 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	N/A	-

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

Contenido	Valoración	Puntuación
2.11.2.2.1. Criterios empleados	No utilizan criterios externos para obtener evidencias de este tipo de validez	
2.11.2.2.2. Calidad de los criterios empleados		
2.11.2.2.3. Relación temporal entre test y criterio		
2.11.2.2.4. Tamaño de las muestras		
2.11.2.2.5. Procedimiento de selección de las muestras		
2.11.2.2.6. Promedio de las correlaciones del test con los criterios		

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Adecuada, pero con algunas carencias- Adecuada	2,5
2.11.3.2. Funcionamiento diferencial de los ítems	Adecuado	3

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:

Las evidencias de validez presentadas en el manual impreso de la Batería EBE proporcionan información sobre diversas evidencias de validez: de contenido, basada en la estructura interna y validez convergente. En términos generales, las fuentes de validación utilizadas son adecuadas, aunque presentan algunas limitaciones, que es necesario tener presentes para interpretar correctamente los datos presentados.

Validez de contenido

- En cuanto a la validez basada en el contenido, se consultó a siete expertos académicos e investigadores sobre la claridad, coherencia, relevancia y suficiencia de los ítems de cada dimensión y escala. Se informa de un grado de acuerdo global del 80% en cada escala, pero no se presentan estadísticos para valorar estos 4 aspectos ni estimaciones de la variabilidad en los mismos.
- El índice de homogeneidad de un ítem (la correlación elemento-total corregida) no es el índice adecuado para justificar que todos los ítems son representativos de un dominio. Los índices de homogeneidad muestran la tendencia de los ítems a estar intensamente relacionados, pero que correlacionen muy alto no garantiza que sean representativos del dominio. Estos análisis de homogeneidad de los ítems, junto con otras evidencias, suelen formar parte de un epígrafe independiente de “análisis de ítems” y es interesante informar de ellos en el manual.
- Con todo, lo más relevante es que los autores en el manual dicen realizar un análisis de diferencias de medias en los ítems en función de sexo mediante una prueba U de Mann-Whitney para evaluar el funcionamiento diferencial de los ítems (DIF). En primer lugar, al igual que en el caso anterior, el epígrafe de validez de contenido no parece ser el más adecuado para la inclusión de análisis de DIF. Y en segundo lugar, y, mucho más importante, con la prueba empleada no es posible evaluar la presencia funcionamiento diferencial de los ítems (DIF) respecto al sexo. Esto se debe a que la técnica de análisis empleada no permite distinguir entre el impacto del sexo y el DIF propiamente dicho. Una vez resaltada la valoración de este aspecto basada en la información presente en el Manual, es conveniente señalar que la Editorial dispone de unos nuevos análisis sobre esta cuestión (que fueron compartidos con el coordinador) que ya sí resultan adecuados y que sería conveniente que incluyeran en futuras ediciones del Manual o que pusieran a disposición de cualquier usuario interesado.
- Adicionalmente se recomienda revisar algunos de los valores reportados en esta sucesión de tablas pues son incorrectos (i.e. tabla 5-19, pag.43: se interpretan como diferencias significativas aquellas que tienen un nivel de significación superior a 0,05 y viceversa).

Estructura interna

- La estructura interna de las escalas fue analizada mediante sucesivos análisis factoriales exploratorios y confirmatorios. Conviene comenzar resaltando que la justificación de método de extracción y del método de rotación son insuficientes, y además, incongruentes con información facilitada previamente. Así, si se concluye que ninguno de los ítems tiene una distribución normal (pag. 38) el método de extracción de Máxima Verosimilitud no es el más recomendable. Sería deseable emplear métodos de extracción más robustos.
- Respecto a las evidencias de la estructura interna para la Escala de Evaluación de la Regulación Interna y Externa del Aprendizaje, se extrajeron seis factores ortogonales, los cuales se agruparon en dos subescalas de 18 ítems cada una: Regulación Personal y Regulación Contextual. Esta estructura fue evaluada posteriormente mediante un análisis factorial confirmatorio. Sin embargo, no queda claro si se probaron otros modelos confirmatorios distintos al obtenido a partir del modelo factorial exploratorio. Tampoco se especifica si se realizó un estudio de validación cruzada o si las respuestas de los mismos participantes fueron utilizadas en ambos análisis. Esta falta de claridad podría

comprometer la validez de los resultados obtenidos. En este sentido, se recomienda también revisar los grados de libertad reportados en la Tabla 5-21 para garantizar la consistencia de los análisis.

- En cuanto a la Escala Abreviada para la Evaluación Interactiva del Proceso de Enseñanza-Aprendizaje, se extrajeron diez factores ortogonales, que luego se agruparon en dos subescalas de 27 y 32 ítems: Enseñanza Reguladora y Aprendizaje Autorregulado. Al igual que en el caso anterior, esta estructura fue analizada mediante un modelo factorial confirmatorio. Sin embargo, nuevamente no se aclara si se probaron otros modelos confirmatorios, ni si se realizó una validación cruzada, lo que podría comprometer la robustez de los resultados obtenidos.
- Por otro lado, la estructura interna de la Escala de Salud Académica (compuesta por 8 ítems) se configuró en dos factores: salud física (5 ítems) y salud psicológica (3 ítems). Sin embargo, los resultados reportados en el texto no son consistentes con los indicados en la Tabla 5-25. Mientras que el texto señala que estos dos factores explican un 43,5% de la varianza, la tabla reporta un 64,35%. En este caso, se recomienda revisar la subescala de Salud Psicológica y aportar evidencia adicional sobre la representación adecuada del constructo de salud psicológica en el contexto educativo.

Validez convergente

- En este análisis se ha incluido una matriz de correlaciones entre cuatro subescalas (RP, RC, ER y AA) y la escala (SALUD) de la Batería con dos instrumentos: 1) La Escala de Bienestar Psicológico (Díaz et al., 2006) que evalúa 6 dimensiones y 2) La Escala de Flourishing (adaptada por Pozo-Muñoz et al., 2016). Llama la atención porque los autores consideran las subescalas de las dos primeras escalas y, sin embargo, para la tercera el nivel de análisis pasa a ser el de escala ya que no hay ninguna justificación al respecto en el manual. Los resultados muestran una asociación positiva entre las medidas de la Batería EBE y los componentes del bienestar psicológico y el flourishing de los otros instrumentos. Las correlaciones obtenidas fueron generalmente moderadas y estadísticamente significativas, sin que se pueda observar un patrón diferencial o predominante en las correlaciones. Se echa en falta haber incluido en la presente edición del Manual alguna otra evidencia para valorar la validez discriminante, a pesar de que esas evidencias existen. Sería recomendable que las incluyeran en futuras ediciones del Manual o que las pusieran a disposición de cualquier usuario interesado.

Recomendaciones adicionales

- Dado que uno de los objetivos de la batería es evaluar factores relacionados con el bienestar escolar y académico, se sugiere obtener más evidencias sobre la relación de las puntuaciones de la batería con variables contextuales del aprendizaje, como el tipo de centro educativo o las características curriculares. Además, sería valioso investigar las diferencias en las puntuaciones según variables demográficas como el sexo, la edad o el nivel educativo de los participantes.
- Finalmente, sería recomendable incorporar a las futuras ediciones del Manual las nuevas evidencias de las que dispone la Editorial sobre la capacidad predictiva de la batería, es decir, su habilidad para pronosticar el rendimiento académico de los estudiantes. Esto podría ofrecer una perspectiva adicional sobre la utilidad de la batería en contextos educativos aplicados.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Coeficientes de fiabilidad (alfa de Cronbach para escalas, subescalas y factores; y Dos Mitades solo para las escalas). No se reporta fiabilidad del Índice de Bienestar Educativo (IBE)	

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	Un estudio con una muestra grande ($N > 500$)	3
2.12.3.2. Coeficientes de consistencia interna presentados	Alfa de Cronbach y Dos Mitades	
2.12.3.3. Promedio de coeficientes de consistencia	Buena ($0.80 \leq r < 0.85$)	4

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	N/A	
2.12.6.2. Promedio de los coeficientes	N/A	-

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

El estudio de la fiabilidad de las escalas, subescalas y factores de la Batería EBE se ha centrado en la obtención de índices de consistencia interna mediante el alfa de Cronbach. Para el caso de las escalas se ha proporcionado también la estimación dos mitades (coeficiente de Spearman-Brown y de Guttman). Estos índices se han calculado en la muestra total de normalización completa que se compone de estudiantes de Bachillerato y universitarios, las edades están comprendidas entre 17 y 24 años, con una edad media de 22.9 años y una desviación típica de 6.2. El número total de participantes fue de 1268 y la mayoría (68%) fueron mujeres. Para la Escala para la Evaluación de la Regulación Interna y Externa del Aprendizaje (SRL-ERL) los valores de estos coeficientes de fiabilidad se encuentran todos en unos rangos aceptables, con un mínimo de 0,73 y un máximo de 0,93. Para la Escala Abreviada para la evaluación interactiva del proceso de Enseñanza-Aprendizaje (EIPEA-Abreviada) los índices de fiabilidad reportados son también aceptables, con un mínimo de 0,75 y un máximo de 0,96.

Sin embargo, se han detectado varias limitaciones en cuanto a las evidencias de fiabilidad informadas en el Manual.

- Para la Escala de Salud estos valores fueron solamente adecuados. En concreto, la estimación de fiabilidad como consistencia interna para la subescala de Salud Psicológica fue 0,61, siendo una subescala formada por solo 3 ítems. Además, los índices de fiabilidad dos mitades son bastante pobres 0,48. Es cierto que esta escala (y sus dos subescalas) se componen de un menor número de elementos, pero en estos casos se pueden considerar otras aproximaciones para estimar la fiabilidad de la escala (i.e. calcular la fiabilidad mediante el AFC, el omega de McDonald o estimar la fiabilidad con modelos de TRI).
- Dado que las escalas que componen la Batería EBE proceden de instrumentos desarrollados y validados previamente por los autores, la principal aportación es la posibilidad de obtener el Índice de Bienestar Educativo (IBE) a partir de la información de las tres escalas. Sin embargo, en la versión actual del Manual no se presenta ninguna estimación de la fiabilidad por ningún procedimiento para el IBE. Sería recomendable que incluyeran esta información en futuras ediciones del Manual o que la pusieran a disposición de cualquier usuario interesado.

- Hubiera resultado útil disponer de una tabla distribución de frecuencias de la muestra total empleada según edad, curso y sexo. En concreto, dado que se van a considerar dos subgrupos en función de edad para elaborar las normas sería interesante, al menos para los mismos, haber informado de las fiabilidades en cada uno de ellos.
- Una limitación evidente es la ausencia de información sobre la consistencia entre mediciones realizadas en diferentes momentos temporales. La inclusión de estos datos proporcionaría una evaluación más completa de la estabilidad y consistencia del test a lo largo del tiempo. Adicionalmente, sería beneficioso incluir información sobre fiabilidad basada en la Teoría de Respuesta al Ítem (TRI), específicamente las funciones de información de las escalas y subescalas. Esto ofrecería una visión más detallada sobre la precisión de la medición en diferentes niveles de rasgo.

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, considerando las diferencias entre poblaciones	2
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	Incidental	
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	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:

La Batería EBE ofrece una interpretación normativa (basada en normas).

Los baremos presentados son adecuados, aunque presentan ciertas limitaciones que deberían invitar a los usuarios a interpretarlos con cierta precaución. La muestra utilizada para la baremación está compuesta por estudiantes de Bachillerato y Universitarios, con un rango de edades entre 17 y 24 años, con una edad media de 22,9 años y una desviación estándar de 6,2. El total de participantes fue de 1.268, de los cuales un 68% eran mujeres. No se proporciona información específica sobre las fechas de aplicación de los instrumentos, aunque, basándose en la información disponible en el manual, se puede inferir que la aplicación de los mismos tuvo lugar con posterioridad al 2017. Además, se dice que se trabajará sobre dos baremos (17-20 años y 21-24 años) pero no se informa en el Manual del tamaño de la submuestra del grupo normativo que se empleará para su construcción, aunque se infiere también que deben ser moderados.

No se incluyen detalles adicionales sobre las características sociodemográficas de los participantes, ni sobre el número de ellos por nivel educativo (i.e. Institutos de Secundaria o Universidad) ni sobre los centros educativos (i.e. públicos, privados, concertados) o la diversidad territorial de la muestra. Tampoco se dice explícitamente nada de la utilización del instrumento asociado a contextos formativos de grados medios o superiores. Algunas de estas características tienen un impacto clave que podría justificar la necesidad de elaborar baremos diferenciados en función de los diferentes niveles en las mismas.

Por otra parte, pero en línea con lo anterior, dado que no se proporcionan datos sobre posibles diferencias en función del sexo, no es posible determinar si sería necesario establecer baremos diferenciados para esta variable. No obstante, sería recomendable realizar un análisis más exhaustivo sobre esta cuestión.

Cuando se describe la construcción de las normas se especifica que se calcularon basados en los Percentiles y las Puntuaciones T. Sin embargo, en el Manual no se presentan las tablas con estas normas. Lo único que se especifica en el apartado "Construcción de las normas" es lo siguiente: *"las normas se elaboraron ajustando la distribución de frecuencias obtenidas para cada escala a puntuaciones estándar de probabilidad normal utilizando los rangos percentiles asociados a las puntuaciones brutas de cada escala"*. En el informe aparecen las puntuaciones directas y las puntuaciones T. El perfil gráfico se establece basado en las puntuaciones percentiles (con sus tres niveles delimitados: bajo, medio y alto).

Finalmente, se observa una incongruencia en la información proporcionada en la documentación entre el rango de edad recomendado para la aplicación de la batería (de 17 a 21 años), la media de edad de los participantes en la que se basan los datos de baremación (22,9 años) y la construcción de dos baremos diferenciados según los grupos de edad (de 17 a 20 años y de 21 a 24 años).