

Evaluación de

DABS. Escala de

Diagnóstico de

Conducta Adaptativa

 **DABS** 

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	DABS. Escala de Diagnóstico de Conducta Adaptativa.
Autor	Marc J. Tassé, Robert L. Schalock, Giulia Balboni, Henry (Hank) Bersani, Jr., Sharon A. Borthwick-Duffy, Scott Spreat, David Thissen, Keith F. Widaman & Dalun Zhang.
Autor de la adaptación española	Miguel Ángel Verdugo Alonso, Benito Arias Martínez y Patricia Navas Macho
Variable(s)	Conducta adaptativa
Áreas de aplicación	Psicología Clínica Psicología del Desarrollo Servicios Sociales Inteligencia
Soporte	Administración oral (papel y lápiz). El test se realiza mediante entrevista estructurada a un adulto que conozca al evaluado/a desde hace al menos 6 meses. La corrección es electrónica (a través de TEAcorrige).

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Excelente	5
Fundamentación teórica	Excelente	5
Adaptación	Excelente	5
Análisis de ítems	Buena-Excelente	4.5
Validez: contenido	Excelente	5
Validez: relación con otras variables	Adecuada-Buena	3.7
Validez: estructura interna	Buena-Excelente	4.5
Validez: análisis del DIF	Adecuado	3
Fiabilidad: equivalencia	-	-
Fiabilidad: consistencia interna	Buena-Excelente	4.5
Fiabilidad: estabilidad	Adecuada	3
Fiabilidad: TRI	Buena	4
Fiabilidad inter-jueces	Excelente	5
Baremos e interpretación de puntuaciones	Buena-Excelente	4.3

Comentarios generales

La principal aportación de la "DABS. Escala de Diagnóstico de Conducta Adaptativa" es la sistematización de la evaluación de la conducta adaptativa. Presentar dificultades a la hora de realizar conductas esperables según la edad es uno de los principales criterios que definen la discapacidad intelectual. A pesar de su importancia, el número de instrumentos que existe para evaluar este constructo es escaso.

La DABS es una adaptación de una prueba internacional desarrollada bajo el auspicio de la asociación de referencia en EEUU sobre discapacidad intelectual, con el objetivo de superar las limitaciones de las pruebas previas e incorporar mejoras teóricas y métricas de los últimos años. En España la adaptación ha sido llevada a cabo por uno de los equipos de referencia (nacional e internacional) sobre discapacidad. Para el desarrollo del instrumento se realiza una revisión exhaustiva de instrumentos previos y desarrollos teóricos en el diagnóstico, se cuenta con profesionales y expertos y muestras –dadas las características de la población objetivo de la prueba– que podemos considerar como amplias. Se aplican métodos psicométricos rigurosos y se documenta la capacidad de la prueba para ofrecer un diagnóstico con una precisión altamente satisfactoria. Se ha tenido mucho cuidado en la adaptación al castellano: se ha partido de un amplio banco de ítems, se han realizado estudios piloto, se ha contado con un proceso de retro-traducción y con la presencia de jueces expertos en la temática para valorar el contenido de los ítems. Así mismo se incluyen en los ítems diferentes notas para poder adaptarlo a diferentes entornos culturales o ejemplos que aclaran el contenido de la información. En este sentido, puede considerarse una prueba ejemplar en su desarrollo. Además, esta prueba supera algunas de las deficiencias de instrumentos similares adaptados al castellano o de procedencia anglosajona (p.ej., baremos actualizados, cálculo de las puntuaciones según procedimientos de TRI, definición operativa de las variables a medir).

La prueba tiene tres versiones según tres grupos de edad: 4 a 8 años, 9 a 15 años y 16 a 21 años. Cada versión consta con 75 ítems (25 para cada una de las tres subescalas: habilidades conceptuales, sociales y prácticas). Los ítems han sido elegidos teniendo en cuenta la función de información de cada uno de ellos en torno al punto de corte para el diagnóstico (2 desviaciones típicas por debajo de la media; $CI < 70$), para ello se ha utilizado el modelo de respuesta graduada (estimaciones de la Teoría de la Respuesta al Ítem, TRI). La aplicación de la prueba se realiza mediante entrevista estructurada a una persona que conozca al evaluado/a durante al menos seis meses.

El uso de la TRI es una de las fortalezas de la prueba y proporciona varias ventajas. Una ventaja es que optimiza la selección de los ítems, como ya se ha mencionado. Otra ventaja es que, a través de un diseño de anclaje, se obtienen las puntuaciones de los evaluados que responden a distintas versiones en la misma métrica. Una tercera ventaja es que se facilita el tratamiento de las respuestas NA (no aplicables), ya que la puntuación es el nivel de rasgo esperado según las respuestas dadas y el error típico tiene en cuenta las respuestas proporcionadas.

Las propiedades psicométricas de la prueba son, en general, buenas o excelentes como demuestran los valores de consistencia interna, las evidencias de validez de estructura y de criterio (diagnóstico y edad). El apartado de evidencias de validez de criterio diagnóstico es el más amplio de la prueba, presentando numerosos índices para asegurarse de que las puntuaciones obtenidas son coherentes con el diagnóstico previo. Los resultados muestran una alta especificidad (mayor del 95% en la mayoría de los grupos y versiones) y una elevada sensibilidad (próxima o superior al 70% para Habilidades sociales -Hs- y Habilidades prácticas -Hp-; próxima o superior al 85% para Hc y Conducta adaptativa -CA-). A nivel global, se informa que el 92% de las personas con discapacidad intelectual (DI) estarían bien clasificadas, por tener puntuaciones por debajo del punto de corte en una o más de las escalas DABS.

La prueba también adolece de algunas limitaciones, dentro de sus muchos méritos. Algunos aspectos mejorables son los estudios que se han realizado para la generación de evidencias de fiabilidad test-retest (estabilidad temporal), la fiabilidad Interjueces o las evidencias de validez con otras pruebas. Estos estudios comparten el problema de haberse realizado con muestras incidentales de tamaño muy reducido y, por lo tanto, los resultados pueden ser inestables.

Con relación a las escalas específicas, debería profundizarse en la validez discriminante de algunas de ellas, dada la alta correlación entre dimensiones que se encuentra en los grupos de mayor edad (> 9 años), y también en la menor fiabilidad interjueces para la escala de habilidades prácticas.

Con relación a la muestra cabe destacar que se ha realizado un muestreo de conveniencia y se ha sobremuestreado a las personas con discapacidad intelectual leve. Esto es un

valor, ya que permite ilustrar la capacidad discriminativa de la prueba en situaciones de discapacidad leve.

Por último, puesto que se ha aplicado un modelo de TRI, sería conveniente disponer de información sobre precisión condicionada al nivel de rasgo, para saber en qué medida el diagnóstico es más o menos confiable para niveles próximos o alejados al punto de corte. En el manual se indica que se han seleccionado aquellos ítems que proporcionan la máxima información (y, en consecuencia, el menor error de medida) en niveles próximos al punto de corte (es decir, $\approx -2D_t$).

Asimismo, podría incorporarse información más detallada sobre algunos análisis (p.ej., resultados de los modelos de análisis factorial exploratorio y análisis de funcionamiento diferencial de los ítems que han llevado a clasificar algunos ítems de anclaje como "no enlazables"). También sería importante proporcionar mayor justificación para el diseño de anclaje, ya que para la escala de Habilidades conceptuales en el grupo de menor edad (4 a 8 años) se dispone de únicamente 4 ítems de anclaje invariantes. También podría valorarse la inclusión de más información sobre los estudios de invarianza de las escalas a través de género.

En conclusión, la DABS se puede considerar una prueba útil para el ámbito clínico y educativo, al permitir evaluar uno de los criterios de discapacidad intelectual. Puesto que el número de instrumentos que existe para evaluar la conducta adaptativa es escaso, parece claro que la prueba viene a cubrir un hueco importante, que es de relevancia social.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

DABS. Escala de Diagnóstico de Conducta Adaptativa.

1.2. Nombre del test en su versión original:

DABS. Diagnostic Adaptive Behavior Scale (DABS).

1.3. Autor/es del test original:

Marc J. Tassé, Robert L. Schalock, Giulia Balboni, Henry (Hank) Bersani, Jr., Sharon A. Borthwick-Duffy, Scott Spreat, David Thissen, Keith F. Widaman & Dalun Zhang.

1.4. Autor/es de la adaptación española:

Miguel Ángel Verdugo Alonso, Benito Arias Martínez y Patricia Navas Macho.

1.5. Editor del test en su versión original:

American Association on Intellectual and Developmental Disabilities

1.6. Editor de la adaptación española:

Hogrefe TEA Ediciones

1.7. Fecha de publicación del test original:

2017

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

2021

1.9. Fecha de la última revisión del test:

--

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

Inteligencia, Habilidades, Escalas de desarrollo, Escalas clínicas

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

La DABS (Escala de Diagnóstico de Conducta Adaptativa) evalúa la existencia de limitaciones significativas en la conducta adaptativa que, junto con una puntuación de inteligencia inferior a dos desviaciones típicas ($CI < 70$) y el inicio de las limitaciones durante el periodo de desarrollo, son los tres criterios principales para el diagnóstico de la discapacidad intelectual (DI). Se entiende por conducta adaptativa el conjunto de habilidades conceptuales, sociales y prácticas que han debido ser aprendidas a lo largo de la vida para manejarse en las actividades cotidianas. La prueba ofrece cuatro puntuaciones de carácter clínico: una puntuación global (CA) y tres puntuaciones para cada una de las habilidades que componen la conducta adaptativa: habilidades conceptuales, habilidades sociales y habilidades prácticas.

HABILIDADES CONCEPTUALES (Hc). Las habilidades conceptuales son aquellas que reflejan la aplicación de competencias académicas, previamente adquiridas, a situaciones de la vida cotidiana. Estas habilidades incluyen: manejo del lenguaje, habilidades lectoescritoras, autodirección, conceptos temporales, conceptos numéricos y métricos, aspectos relativos a la comprensión del dinero y solución de problemas.

HABILIDADES SOCIALES (Hs). Las habilidades sociales son las que permiten interactuar con otras personas con éxito. Estas habilidades incluyen: relaciones interpersonales, responsabilidad, autoestima, prudencia, ingenuidad, modales, cumplimiento de normas y resolución de problemas sociales.

HABILIDADES PRÁCTICAS (Hp). Las habilidades prácticas son las básicas e instrumentales de la vida diaria. Estas habilidades incluyen: actividades de la vida diaria, habilidades ocupacionales, manejo del dinero, seguridad, cuidado de la salud, realización de viajes y desplazamientos, capacidad de seguir rutinas diarias y uso del teléfono.

El rango de edades para las personas objeto de evaluación es desde los 4 hasta los 21 años. La DABS tiene tres versiones según el rango de edad en las que varía el contenido de las preguntas y las habilidades específicas evaluadas, pero no las áreas de evaluación (i.e., habilidades conceptuales, sociales y prácticas). En concreto, los tres rangos de edad propuestos son: 4 a 8 años, 9 a 15 años y 16 a 21 años.

La escala es un instrumento de aplicación individual, vía entrevista estructurada a partir de los ítems de la prueba a informantes clave, que conozcan bien a la persona objeto de evaluación (un mínimo de 6 meses). El formato de respuesta es de categorías ordenadas (nunca o casi nunca lo hace, lo hace con ayuda o cuando se lo recuerdan, lo hace a veces sin ayuda, siempre o casi siempre lo hace sin ayuda). En aquellos casos en los que el informante no disponga de conocimiento exacto o se considere que el entorno puede haber limitado la expresión de la conducta, los ítems pueden puntuarse como NA (No aplicable).

La prueba evalúa no lo que la persona evaluada puede llegar a hacer sino lo que habitualmente realiza. Además, se tienen en cuenta las oportunidades de las que disponga, que pueden afectar a la expresión de las conductas. El objetivo principal es proporcionar un diagnóstico global, más que de detectar puntos fuertes y limitaciones del evaluado, con relación a habilidades concretas específicas. Las personas interesadas en detectar las fortalezas y debilidades del evaluado deberán complementar la evaluación de la DABS con otras escalas destinadas a ese fin.

Para la construcción de la prueba se ha utilizado el modelo de la Teoría de la Respuesta al Ítem (TRI); en concreto, el modelo de respuesta graduada de Samejima (1969). La DABS ofrece puntuaciones en escala CI ($M = 100$, $DT = 15$) para cada una de las tres áreas (Hc, Hs y Hp), y una puntuación general (CA), siendo necesaria una puntuación inferior a 70 en una de las áreas, o en la puntuación global, para poder diagnosticar como DI.

La corrección es electrónica y se realiza a través de la plataforma de TEACorrige.

1.12. Áreas de aplicación:

Psicología clínica, Psicología educativa, Servicios sociales.

1.13. Formato de los ítems:

Respuesta graduada / Tipo Likert.

1.14. Número de ítems:

Excluyendo los ítems de identificación previos, las tres versiones (adaptadas según la edad) tienen un

total de 75 ítems, 25 ítems para cada escala: Habilidades conceptuales, Habilidades sociales y Habilidades prácticas. Existe una versión distinta para cada grupo de edad (4 a 8 años, 9 a 15 años y 16 a 21 años).

1.15. Soporte:

Administración oral, Papel y lápiz.

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

Nivel C. La persona a cargo de la evaluación ha de ser un profesional (p.ej., psicólogo u orientador escolar) con experiencia en el ámbito de la evaluación psicológica y en el campo de la discapacidad intelectual o del desarrollo. Se indica que, de manera excepcional, la escala puede ser aplicada por personas sin esa cualificación siempre que cuenten con amplia experiencia en la realización de evaluaciones conductuales o dominen la evaluación psicológica.

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

La DABS está destinada a evaluar personas en el rango de edad de 4 a 21 años. Se entiende que los niveles de conducta adaptativa ya se han desarrollado pasados los 21 años, por lo que podría usarse pasada esta edad. Ahora bien, se advierte de que la prueba no está baremada ni analizada pasada esa franja de edad, por lo que cualquier interpretación tendría que realizarse con cautela.

La DABS es una escala orientada al diagnóstico, que permite la evaluación de uno de los criterios diagnósticos de la discapacidad intelectual. Por tanto, tiene objetivos clínicos y la población de aplicación natural es la de personas para las que sea necesario realizar ese diagnóstico.

La DABS no es un autoinforme sino que se aplica mediante entrevista a personas que conozcan en profundidad a la persona objeto de evaluación (p.ej., familiares, amigos, docentes). La DABS puede ser completada por varios informantes, lo que permite obtener información en distintos contextos y desde distintos puntos de vista, si bien es tarea del evaluador resolver las discrepancias que puedan aparecer.

Se indica que la DABS incluye ítems que minimizan la posible influencia en la evaluación de factores culturales o geográficos. Además, en algunos ítems se incluyen aclaraciones que indican la posibilidad de adaptar algunos términos (p.ej., en preguntas relacionadas con el tipo de moneda). Por último, en el manual se destaca la necesidad de tener en cuenta el entorno cultural (p.ej., que algunas conductas puedan ser especialmente importantes en una cultura concreta o que la persona evaluada haya dispuesto de suficiente tiempo para adaptarse al entorno cultural en el que se encuentra). En aquellos casos en los que se considere que el entorno puede haber limitado la expresión de la conducta, el ítem debería puntuarse como NA (No aplicable).

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

El test tiene tres versiones en las que cambia el contenido de los ítems, pero no los tres tipos de habilidades evaluadas (conceptuales, sociales y prácticas). Las tres versiones hacen referencia a diferentes rangos de edad: 4 a 8 años, 9 a 15 años y 16 a 21 años.

1.19. Procedimiento de corrección:

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

1.20. Puntuaciones:

Las puntuaciones solo pueden obtenerse mediante corrección informática al ser puntuaciones derivadas de la Teoría de Respuesta al Ítem (TRI; en concreto, el modelo de Respuesta Graduada). En concreto, se realizan estimaciones bayesianas (EAP) del nivel de rasgo, para las puntuaciones en las tres subescalas (Hc, Hs y Hp). El índice de Conducta Adaptativa (CA), se obtiene como un promedio de las anteriores.

Estas puntuaciones son transformadas linealmente a una escala CI ($M = 100$, $DT = 15$), tomando como referencia la población sin discapacidad intelectual de la edad del evaluado. No se realiza ningún tipo de corrección por azar o inversión de los ítems, puesto que el objetivo del test es la evaluación del rendimiento típico (habitual) de la persona y el formato de aplicación es vía entrevista estructurada a un informante. El uso de la TRI permite la presencia de omisiones (p.ej., si un ítem se considera no aplicable), aunque la recomendación es que se conteste a todos los ítems ya que cuanto mayor sea el número de omisiones mayor será el error típico de medida.

1.21. Escalas utilizadas:

Puntuaciones estandarizadas. Escala de CI (Media 100 y desviación típica 15).

1.22. Posibilidad de obtener informes automatizados:

Sí. A través de la corrección automática, se obtienen gráficos de perfil, sencillos, que incluyen las puntuaciones Hc, Hs, Hp y CA, en escala de CI, con un intervalo de confianza del 95%. No obstante, el perfil no incluye texto que ayude a interpretar las puntuaciones.

1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: 30 minutos aproximadamente.

En aplicación colectiva: Dado el modo de aplicación (entrevista), no es viable la aplicación colectiva.

1.24. Documentación aportada por el editor:

Manual.

1.25. Precio de un juego completo de la prueba:

97,10 € –Juego completo. (Manual, 10 Ejemplares de cada tipo, Pin de corrección 30 usos). Fecha de consulta: 09/04/2022.

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

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

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

29,77 € –Kit corrección (25 Ejemplares, Pin 25 usos). Fecha de consulta: 09/04/2022.

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Excelente	5
2.2. Calidad de la documentación aportada	Excelente	5
2.3. Fundamentación teórica	Excelente	5
2.4. Adaptación del test	Excelente	5
2.5. Desarrollo de los ítems del test	Excelente	5
2.6. Calidad de las instrucciones para el participante	Excelente	5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Excelente	5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Excelente	5
2.10. Datos sobre el análisis de los ítems	Buenos-Excelentes	4.5

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	Excelente	5
2.11.1.2. Consultas a expertos	Se ha consultado a un amplio número de expertos mediante un procedimiento sistematizado (N > 30)	5

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

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

Contenido	Valoración	Puntuación
2.11.2.1.1. Diseños empleados	Correlaciones con otros test.	
2.11.2.1.2. Tamaño de las muestras	Un estudio con una muestra pequeña (N < 200)	1
2.11.2.1.3. Procedimiento de selección de las muestras	<p>Incidental. La muestra general está compuesta por personas con y sin discapacidad intelectual entre los 4 y 21 años con, al menos, dos años de residencia en España. Para su obtención se contó con la colaboración de distintas Direcciones Generales de Planificación, Ordenación e Inspección Educativa y entidades del ámbito de la discapacidad intelectual como Plena Inclusión o Down España. El criterio fundamental era el acceso y disponibilidad de los profesionales de los centros. Se sobremuestra a personas con discapacidad intelectual leve.</p> <p>Para obtener la relación con las puntuaciones en ICAP, la muestra final está compuesta por 32 participantes dentro del grupo de personas con discapacidad intelectual para las que se disponía de valoración en el inventario ICAP (6.1% de la muestra con el diagnóstico).</p>	
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	Excelente ($r \geq 0.70$)	5
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no	-	-

relacionados		
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	-	-
2.11.2.1.8. Resultados de las diferencias intergrupo	-	-

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 criterios empleados son la edad y el diagnóstico. Los criterios utilizados se justifican puesto que la conducta adaptativa es una variable que evoluciona con la edad. No obstante, el criterio más importante es el diagnóstico y la comparación de las puntuaciones de las personas que cumplan los criterios para discapacidad intelectual (población clínica) con personas que no cumplen los criterios diagnósticos (población no clínica).	
2.11.2.2.2. Calidad de los criterios empleados	Excelente	5
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 ($N \geq 200$)	3
2.11.2.2.5. Procedimiento de selección de las muestras	<p>La muestra general está compuesta por personas con y sin discapacidad intelectual entre los 4 y 21 años con, al menos, dos años de residencia en España. Para su obtención se contó con la colaboración de distintas Direcciones Generales de Planificación, Ordenación e Inspección Educativa y entidades del ámbito de la discapacidad intelectual como Plena Inclusión o Down España. El criterio fundamental era el acceso y disponibilidad de los profesionales de los centros. Se sobremuestra a personas con discapacidad intelectual leve.</p> <p>La muestra final estaba dividida entre personas con diagnóstico de discapacidad intelectual (grupo clínico) y personas sin diagnóstico (grupo no clínico). Por lo tanto, los grupos vienen dados según el diagnóstico. Para la relación con el criterio edad, se realizaron regresiones no paramétricas entre las puntuaciones obtenidas con el instrumento y la variable edad. Las muestras se componen de entre 288 y 388 personas según la versión por grupo de edad. También se realizaron análisis de sensibilidad y especificidad con respecto a la clasificación diagnóstica de DI.</p>	
2.11.2.2.6. Promedio de las correlaciones del	Excelente ($r \geq 0.55$)	5

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	Buena-Excelente	4.5
2.11.3.2. Funcionamiento diferencial de los ítems	Adecuados	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	Sí. Se ofrecen algunas indicaciones para la evaluación y puntuación de personas con diversidad funcional. Por ejemplo, indica que, para personas con una discapacidad de tipo físico, debe valorarse a la persona teniendo en cuenta la tecnología disponible al uso (p.ej., ayudas técnicas).	

2.11.5. Comentarios generales sobre evidencias de validez:

Las evidencias de validez se resumen en los siguientes puntos:

1. Evidencias de validez de contenido. El desarrollo de la prueba original se plantea dentro de la Asociación Americana de Discapacidades Intelectuales y del Desarrollo, una sociedad científica de referencia y prestigio internacional. El diseño de la prueba ha recaído, por tanto, en expertos. Se realiza un adecuado trabajo de inspección de propuestas previas para la evaluación del constructo, se cotejan con el marco teórico actual, y se define claramente el dominio que ha de evaluar la prueba. Se toman las medidas requeridas para que los ítems inicialmente desarrollados y finalmente incluidos cubran de forma satisfactoria este dominio. En la adaptación española se parte de un banco inicial de 260 ítems, el banco previo a la selección de ítems en el test original. Se ha seguido el proceso de retro-traducción y se ha contado con jueces expertos en la temática para valorar el contenido de los ítems en idoneidad, importancia y observabilidad.

2. Las evidencias de validez con relación a la estructura interna son satisfactorias. Para cada versión de la escala, se realiza un análisis paralelo que apoya la presencia de tres dimensiones. Se contrastan diferentes modelos confirmatorios y exploratorios (p.ej., un factor, tres factores, bi-factor), utilizando el estimador apropiado (WLSMV) y obteniéndose evidencia que permite mantener el modelo confirmatorio de tres factores correlacionados, con índices de ajuste (p.ej., RSMEA, CFI) excelentes (CFI > 0,96; RMSEA < 0,05) y con saturaciones factoriales aceptables para todos los ítems (> 0,38). Todo ello indica que la estructura propuesta se puede mantener.

Por último, con el fin de comprobar el grado de independencia de cada una de las subescalas que conforman el test, se ha tenido en cuenta la varianza media extraída para cada factor y las correlaciones entre factores. Los resultados son generalmente satisfactorios, si bien las correlaciones entre factores son, en algunos casos, excesivamente elevadas (p.ej., superiores a 0,8 en el grupo de edad de 9 a 15 años). En la misma línea, el modelo de un factor muestra un ajuste excelente en dos de los rangos de edad (9 a 15 años; 16 a 21 años). Aunque se entiende la decisión de los autores de mantener el modelo de tres factores, que se replica a través de los tres grupos de edad, sería recomendable explorar en profundidad esta circunstancia.

Los análisis fueron realizados con muestras de tamaño moderado (4 a 8 años: N = 388; 9 a 15 años: N = 371; 16 a 21 años: N= 288).

3. Evidencias de validez relacionadas con el proceso de respuesta. Se realizan estudios piloto sobre comprensibilidad de los ítems. Se descartan aquellos ítems que generan más problemas para ser respondidos. Se redactan los ítems con información suficiente como para reducir la incertidumbre sobre su comprensión. Se definen claramente las condiciones que ha de reunir el informante para contar con elementos de juicio como para ofrecer respuestas informativas. Se busca que los ítems evalúen aspectos 'visibles', detectables en la conducta de la persona que se busca evaluar. Al anclar las respuestas en elementos observables se intenta reducir elementos de posible distorsión en la respuesta.

4. Evidencias de relación con otras medidas de conducta adaptativa. Este es quizás el aspecto con mayor capacidad de mejora, puesto que, para la segunda prueba se ha utilizado una muestra muy pequeña ($n = 32$; 6.1% de las personas diagnosticadas con discapacidad intelectual). La selección de esta muestra fue totalmente incidental, se aprovechó que ya se disponía de las puntuaciones en el ICAP (Inventario para la Planificación de Servicios y la Programación Individual) de estas personas para realizar los cálculos de correlaciones. Los resultados de las correlaciones son adecuados. La evidencia es favorable y se encuentran correlaciones elevadas (p.ej., de 0,78 entre la puntuación total del ICAP y la puntuación global de la DABS, CA). También se encuentra evidencia de validez convergente para las subescalas específicas (correlaciones de 0,78, entre destrezas de la vida personal y Hp; de 0,79 y 0,61, entre destrezas sociales y comunicativas y Hc y Hs, respectivamente; de 0,66 y 0,65, entre destrezas de la vida en comunidad y Hs y Hp, respectivamente). Aunque el tamaño muestral de este estudio es pequeño, estos resultados van en la misma línea de la validación de la DABS en EE.UU., donde se estudiaron las relaciones con otras medidas de conducta adaptativa, como la Escala Vineland de Conducta Adaptativa - II (Vineland II; Sparrow et al., 2005).

Como propuesta de mejora, considerando la presencia de otros instrumentos en castellano que miden conducta adaptativa (ej., ABAS-II), hubiera sido adecuado realizar un estudio específico para mostrar el grado de convergencia de la DABS con otras medidas previamente publicadas.

5. Evidencias de validez con diferentes criterios: edad. En este caso, no se diferencia por versión aplicada y se trabaja con la muestra completa ($N = 1047$). Ello es posible porque las puntuaciones de TRI se encuentran en la misma métrica (a pesar de que los evaluados de distinta edad han respondido a versiones distintas, estas disponen de ítems de anclaje). Se comprueba que las puntuaciones en la conducta adaptativa aumentan con la edad (aspecto teóricamente relevante). Además, se observa que para el caso de las personas sin discapacidad intelectual se encuentra un efecto techo, es decir, sus puntuaciones se agrupan en la parte superior de la escala. Los ítems han sido seleccionados para ser máximamente informativos en el punto de corte ($CI < 70$), así que este efecto techo es esperable y no se considera problemático.

6. Evidencias de validez con diferentes criterios: diagnóstico. Este es el aspecto más relevante puesto que la prueba tiene un objetivo diagnóstico. En este caso, se trabaja con las puntuaciones de CI separando los análisis para cada versión (4-8 años: $N = 388$; 9-15 años: $N = 371$; 16-21 años: $N = 288$). Para este punto se han realizado numerosos cálculos, de entre los que pueden destacarse las comparaciones de grupos (con diagnóstico y sin diagnóstico de discapacidad intelectual) y la obtención de índices de sensibilidad y especificidad. Con respecto al primer punto, los resultados son excelentes (tamaños del efecto grandes para las diferencias entre grupos). Con respecto al segundo punto los resultados también son buenos pero se presentan algunas dudas relevantes. Más en concreto, los autores definen la sensibilidad como la proporción de personas con condición de discapacidad entre el total de personas que han sido marcadas como casos de discapacidad por la DABS (p. 106). Sin embargo, la sensibilidad es la proporción de evaluados marcados como casos con discapacidad de entre los efectivamente discapacitados. Por tanto, es necesario que los autores revisen y modifiquen la redacción del manual en futuras ediciones.

7. Evidencias sobre el funcionamiento diferencial de los ítems (DIF). En el manual se informa sobre la realización de contrastes sobre la invarianza de los parámetros de los ítems a través de las distintas versiones, aplicadas a grupos con distinto rango de edad. En concreto, se han realizado seis análisis completos del DIF para verificar la invarianza de los parámetros de los ítems de anclaje entre las formas de la DABS, en función de los rangos de edad (tres en la fase de selección de los ítems — calibración inicial— y otros tres en la fase de recalibración de la versión final de la escala). En el manual se indica que se utiliza el programa IRTPRO, que implementa el test de Wald, pero no se informa del procedimiento exacto (p. ej., no se indica si se estableció purificación, que es un paso que resulta conveniente en el análisis del funcionamiento diferencial). Se informa que se utilizó la corrección por comparaciones múltiples de Benjamini-Hochberg, que es un procedimiento adecuado de control de la tasa de falsos positivos. De acuerdo con lo informado en el manual, los resultados mostraron ausencia de DIF en los ítems de enlace.

En general, el proceso de validación es ejemplar y los análisis se realizan con muestras

adecuadas considerando la población clínica a la que se aplica la prueba, si bien, para algunos análisis, sería deseable conseguir muestras de mayor tamaño y representación.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	<p>Varios coeficientes de fiabilidad (para cada escala o subescala).</p> <p>Coefficientes de fiabilidad para diferentes grupos de personas.</p> <p>Error típico de medida para diferentes grupos de personas.</p> <p>Cuantificación del error mediante TRI (Función de información u otros).</p>	

2.12.2. Equivalencia formas paralelas:

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

2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Varias muestras de tamaño moderado	4
2.12.3.2. Coeficientes de consistencia interna presentados	Alfa ordinal y fiabilidad compuesta	
2.12.3.3. Promedio de coeficientes de consistencia	Excelente ($r \geq 0.85$)	5

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Un estudio con una muestra pequeña ($N < 100$)	1
2.12.4.2. Coeficientes de estabilidad	Excelente ($r \geq 0.80$)	5

2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	Un estudio con una muestra grande	3
2.12.5.2. Coeficientes proporcionados	Promedio del Error típico de medida en escala CI.	
2.12.5.3. Tamaño de los coeficientes	Excelente (información ≥ 10)	5

2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	Coefficiente de correlación intraclass (ICC) y coeficientes de correlación de Pearson (r)	
2.12.6.2. Promedio de los coeficientes	Excelente ($r \geq 0.80$)	5

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

Hay varias fortalezas a destacar en el desarrollo del test para obtener una adecuada fiabilidad para las puntuaciones. En primer lugar, se parte de un banco amplio de ítems lo que, tras calibración con TRI, permite seleccionar aquellos más informativos en el punto de corte para el diagnóstico. En segundo lugar, los autores emplean distintas aproximaciones a la valoración de la fiabilidad (consistencia interna, estabilidad, fiabilidad interjueces), lo que también debe ser valorado muy positivamente. Los indicadores de fiabilidad obtenidos son en su mayoría considerados excelentes.

1) Consistencia interna. Con respecto a los índices aportados, el manual presenta datos de consistencia interna (alfa ordinal) excelentes para cada grupo de edad ($> 0,97$). Los resultados de fiabilidad compuesta van en la misma línea, informando de una consistencia interna excelente ($> ,0,95$). Atendiendo también a los errores típicos proporcionados en el manual, derivados a partir de la TRI, se puede comprobar que la consistencia interna es muy elevada.

Los resultados de consistencia interna se han calculado para la muestra total ($N = 1047$). Por tanto, para cada versión por separado, se dispone de muestras de tamaño moderado (entre 288 y 388 personas, para los ítems que no sean de anclaje y, si se ha hecho un tratamiento correcto de los valores perdidos, de muestras de más de 500 personas para los ítems de enlace). Hay que tener en consideración que se trata de un test clínico y, por lo tanto, el acceso a la muestra es más costoso.

2) Test-retest. En cuanto a la estabilidad test-retest, se informa de esta para un intervalo de tiempo medio entre primera y segunda aplicación de 23,3 días (el intervalo de tiempo entre el test y el retest es algo amplio, de 7 a 62 días). Aunque es buena o excelente para todas las escalas ($r > 0,76$, promedio de 0,90), se ha calculado únicamente con una muestra de 37 personas. Por ello, sería recomendable que los adaptadores realizaran estudios con un tamaño muestral superior. Se proporcionan los resultados por escala, pero convendría, además, presentar resultados para cada una de las versiones de la prueba (grupos de edad). Otra limitación es que el intervalo de tiempo es demasiado variable y, en algunos casos, demasiado corto (7 días). Por último, hubiera sido apropiado informar de la estabilidad del diagnóstico a partir de las puntuaciones (no sólo de las puntuaciones).

3) Concordancia entre informantes. El manual aporta índices de concordancia entre diferentes informantes ($N = 21$) con valores de correlación intraclass (ICC) entre adecuados y excelentes ($> 0,66$, promedio de 0,81). Para las escalas Hc y Hs pueden considerarse excelentes ($> 0,91$). Este aspecto es clave, puesto que en el manual se recomienda que el psicólogo clínico administre la prueba a tantos informantes como considere necesarios para evaluar el funcionamiento del evaluado en diferentes contextos (p.ej., casa, escuela, etc.). De nuevo, el tamaño muestral puede ser insuficiente, no se aportan los datos segregados según la versión de la prueba y no se informa de la concordancia en el diagnóstico.

En todo caso, debe valorarse positivamente el que se proporcione este tipo de indicador ya que el informante constituye una fuente adicional que contribuye al error de medida.

4) Por último, el manual presenta estimaciones promedio del error típico de medida (en escala CI) obtenidos con la muestra que se ha utilizado para calibrar los ítems según el modelo de respuesta graduada de Samejima. Se ha utilizado una muestra grande (N = 1047), si bien, dado el diseño de anclaje, la muestra disponible para la calibración de los ítems es menor (p.ej., para los ítems únicos en la muestra de 16 a 21 años, N = 288). Para cada escala se comprobaron los supuestos necesarios para la aplicación de la TRI: a) unidimensionalidad o presencia de un factor dominante, que explicaba entre el 85,4% y el 97,3% de la varianza común; b) Independencia local, encontrando un cumplimiento razonable (< 4% de parejas de ítems con valores altos). Los autores no presentan los datos de la función de información, pero en una nota aclaratoria indican que pueden ser facilitados para el lector interesado (y, en todo caso, pueden deducirse elevados dados los valores de los errores típicos promedio facilitados en la tabla 5.17). No obstante, puesto que se aplica la TRI, hubiera sido recomendable informar del error típico condicionado a distintos niveles de conducta adaptativa.

Cabe destacar que las puntuaciones, tal y como se generan desde la plataforma de corrección, incorporan estimaciones de la incertidumbre a través de los intervalos de confianza y teniendo en cuenta las respuestas disponibles (i.e., excluyendo los NA o respuesta "No Aplicable").

2.13. Baremación e interpretación de las puntuaciones:

2.13.1. Interpretación normativa de las puntuaciones:

Contenido	Valoración	Puntuación
2.13.1.1. Calidad de las normas	Varios baremos dirigidos a diversos estratos poblacionales	4
2.13.1.2. Tamaño de las muestras	Moderado ($300 \leq N < 600$)	3
2.13.1.3. Aplicación de estrategia de tipificación continua	Sí	
2.12.1.4. Procedimiento de selección de las muestras	<p>Incidental: Muestreo no probabilístico incidental (muestra de conveniencia). Se requieren personas con y sin discapacidad intelectual entre los 4 y 21 años con, al menos, dos años de residencia en España. Se solicita y cuenta con la colaboración de distintas Direcciones Generales de Planificación, Ordenación e Inspección Educativa y entidades del ámbito de la discapacidad intelectual como Plena Inclusión o Down España.</p> <p>El criterio fundamental fue el acceso y disponibilidad de los profesionales de los centros que ofrecían algún servicio a personas con y sin discapacidad intelectual (DI). Se decide sobremuestrear a personas con discapacidad intelectual leve. La muestra final tiene un tamaño muestral de 1047 personas (522 con discapacidad). Con la muestra total se estima el modelo de respuesta graduada, pero el baremo final se construye con las personas sin discapacidad ($N = 525$).</p>	
2.12.1.5. Actualización de baremos	Excelente (menos de 10 años)	5

2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	Excelente (Se proporciona evidencia empírica con estudios de calidad que relacionan el punto de corte con un criterio externo, para avalar la adecuación y utilidad de los puntos de corte establecidos)	5

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:

Se parte de una muestra recogida a través de muestreo no probabilístico incidental (muestra de conveniencia). Se requieren personas con y sin discapacidad intelectual entre los 4 y 21 años con, al menos, dos años de residencia en España. Se solicita y cuenta con la colaboración de distintas Direcciones Generales de Planificación, Ordenación e Inspección Educativa y entidades del ámbito de la discapacidad intelectual como Plena Inclusión o Down España. El criterio fundamental fue el acceso y disponibilidad de los profesionales de los centros que ofrecían algún servicio a personas con y sin discapacidad intelectual (DI). Se contactó con el objetivo de valorar a 2050 personas, siendo finalmente evaluadas 1129 (por tanto, la tasa de respuesta del 51.1%). Se decide sobremuestrear a personas con discapacidad intelectual leve. La muestra final, tras considerar las evaluaciones válidas, tiene un tamaño muestral de 1047 personas (522 con discapacidad). En términos generales, el porcentaje de varones era ligeramente mayor (56%) y la representación geográfica limitada (p.ej., 37.2% con residencia en Castilla y León).

Para el cálculo de las puntuaciones de CI ($M = 100$, $DT = 15$) se parte, en primer lugar, de las estimaciones de TRI (por el método EAP) para cada subescala (Hs, Hc y Hp), habiendo calibrado cada dimensión por separado en la muestra completa ($N = 1047$). Asimismo, la puntuación global (CA) se calcula como promedio de las puntuaciones de TRI en las tres dimensiones. Finalmente, las puntuaciones de TRI se transforman linealmente a puntuaciones en escala de CI, tomando como referencia el grupo de personas de la misma edad sin diagnóstico de CI (se establecen intervalos de un año de edad). De cada grupo de edad se tienen en cuenta la media y la desviación típica suavizadas mediante regresión local robusta, lo que contribuye a una mayor precisión en los baremos. En todo caso, puesto que los baremos finales de CI se obtienen tomando las personas sin diagnóstico como grupo de referencia ($N = 525$; 52% varones), el tamaño de la muestra puede considerarse moderado.

Cabe destacar que en los perfiles obtenidos se informa de los intervalos de confianza, que, como se ha mencionado anteriormente, son adecuados. Con relación al punto de corte de 70, este se considera justificado, ya que viene dado en parte por los manuales diagnósticos (p. ej., DSM) y es un criterio puramente estadístico (puntuación inferior a dos desviaciones típicas por debajo de la media). Por lo tanto, no se ha necesitado la presencia de jueces externos para fijarlo. Además, se informa de la sensibilidad y de la especificidad de la prueba, según las cuatro puntuaciones (Hc, Hs, Hp y CA). Por otro lado, aunque el manual informa de la posibilidad de otros puntos de corte alternativos (ver Figura 5.15), la información empírica proporcionada avala el uso del punto de corte de 70 para discriminar personas con y sin diagnóstico de DI.

Dentro de las posibles limitaciones de la representatividad de la muestra, los baremos están razonablemente contruidos, con énfasis en adaptarlos según edad. Pueden señalarse algunas posibles mejoras en futuras versiones del manual:

1) Dado que la prueba está orientada al diagnóstico (estar por debajo del umbral: Sí/No), podrían haberse complementado las puntuaciones numéricas con información directa sobre la probabilidad de estar por debajo del umbral dado el patrón de respuestas.

2) En el manual debería explicitarse que los errores típicos se derivan de los obtenidos para las estimaciones EAP. Asimismo, no acaba de quedar claro cómo se calcula el error típico del Índice de Conducta Adaptativa (p.ej., si se tiene en cuenta o no la correlación entre las dimensiones que se promedian).

3) Por último, a un nivel más técnico, pero con potenciales implicaciones aplicadas, cuando se habla de estimación EAP (expected a posteriori) sería bueno explicitar la distribución a priori empleada, ya que esta puede tener un cierto impacto en los resultados.