



Consejo General de
Colegios Oficiales
de Psicólogos

Miembro de la Federación Europea
de Asociaciones de Psicólogos

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

EVALUACIÓN DE LA BATERÍA

BAS-II

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Características	Descripción
Nombre del test	BAS-II – Escalas de aptitudes intelectuales
Autor	Collin D. Elliott, Pauline Smith y Kay McCullogh
Autor de la adaptación española	David Arribas Águila y Sara Corral Gregorio
Editor de la adaptación española	TEA ediciones
Fecha de la última revisión del test en su adaptación española	2011
Constructo evaluado	Inteligencia, Aptitudes
Área de aplicación	Psicología clínica, Psicología educativa, Neuropsicología y Servicios sociales
Soporte	Lápiz y papel, Manipulativo, Administración oral

Valoración general

Características	Valoración	Puntuación
Calidad de los Materiales y documentación	Excelente	4.5
Fundamentación teórica	Excelente	5
Adaptación Española	Excelente	5
Análisis de los ítems	Excelente	5
Validez de contenido	Buena	4
Validez de constructo	Excelente	5
Análisis del sesgo	Excelente	5
Validez predictiva	Buena	4
Fiabilidad: equivalencia	No se aporta	
Fiabilidad: consistencia interna	Buena	4
Fiabilidad: estabilidad	Adecuada	3
Baremos	Buena	4

Comentarios generales

La batería del BAS-II constituye una excelente elección para la valoración de la inteligencia y del funcionamiento cognitivo en niños y adolescentes. Detrás del test y de la adaptación hay una enorme cantidad de tiempo dedicado y trabajo bien hecho. Goza de unas excelentes propiedades psicométricas y numerosas **fortalezas**, que la convierten en una alternativa eficaz a otras baterías similares. Algunas de ellas se describen a continuación:

1. Cabe destacar su robusta base psicométrica, ya que la aplicación del modelo de Rasch aporta muchas ventajas que no tienen otras baterías. Así, el carácter adaptativo de las pruebas, aplicables a un amplio rango de edad, está sólidamente asentado. De esta forma se le aplican a cada niño aquellos ítems que pueden ser adecuados en dificultad a su edad y al nivel de rendimiento que va manifestando a lo largo del test. El uso de puntos de comienzo, puntos de decisión y criterios de terminación ayuda al psicólogo a proteger la autoestima y la motivación del niño, excluyendo aquellos ítems que puedan resultar extremadamente difíciles o extremadamente fáciles, y permite mantener un balance adecuado entre longitud y precisión del test. Otra ventaja es que el modelo de Rasch permite obtener medidas de precisión de las pruebas en cada nivel de aptitud.
2. Un segundo punto fuerte es su firme base teórica. La composición de la batería se fundamenta en el modelo factorial jerárquico de la Inteligencia de Horn-Cattell y Carroll e incluye medidas de g, Gc, Gf, Gv, Gsm, Glr y Gs. La batería se centra principalmente en los cuatro primeros factores. Así, para cualquier grupo de edad, puede obtenerse un índice general de aptitud (IG). En los niños de menor edad (hasta los tres años y seis meses), las aptitudes no se consideran suficientemente diferenciadas y se obtiene únicamente IG. A partir de esa edad, pueden obtenerse los índices de aptitud verbal (IV), de razonamiento perceptivo (RP) y espacial (IE). A partir de los seis años, el índice de Razonamiento perceptivo (RP) se sustituye por un índice de razonamiento no verbal (RNV).
3. Más allá del modelo factorial, se realiza un esfuerzo importante en describir los procesos cognitivos subyacentes a la resolución de cada prueba que incluyen el razonamiento inductivo, la resolución de problemas no verbales, la formulación y comprobación de hipótesis, entre otros muchos. También se recogen diferentes hipótesis alternativas que pueden explicar la obtención de una baja puntuación en los diferentes tests, haciendo énfasis en la contextualización de los resultados.
4. Se aportan estrategias psicométricamente asentadas para establecer comparaciones intra-individuales e identificar los puntos fuertes y débiles del niño. Así, puede determinarse si las puntuaciones en los índices de aptitud intelectual, IV, RNV, RP o IE, difieren significativamente de la puntuación en el índice general, IG. Igualmente, puede valorarse si las puntuaciones en los tests que conforman un índice de aptitud intelectual difieren entre sí. Finalmente, se puede establecer si las puntuaciones obtenidas en los tests de rendimiento (Aritmética y Ortografía) se ajustan a lo esperado según el rendimiento en los tests cognitivos.

5. Tiene un extenso campo de aplicación. En relación al uso del BAS-II en el ámbito profesional, cabe señalar que esta batería resulta apropiada para la evaluación precisa de niños con dificultades de aprendizaje y para la evaluación de altas capacidades o superdotados. Por lo tanto, su uso es adecuado tanto en el ámbito clínico como en el educativo. La exhaustividad en la descripción de los procesos implicados, la convierte también en una buena candidata para el uso en Neuropsicología. El Manual técnico plantea la posibilidad de establecer una evaluación excluyendo los tests más verbales (p.ej., puede obtenerse el índice general no verbal o IGNU). Esto permite aplicar la batería a niños con discapacidad auditiva o verbal y a niños no castellano-parlantes o extranjeros.

A pesar de las mencionadas virtudes, cabe destacar algunas debilidades o **aspectos mejorables** a tener en cuenta en sucesivas ediciones.

1. En la muestra española, algunos tamaños muestrales para los estudios de relaciones entre el BAS-II y otras baterías son escasos (menores de 25 en algunos casos). Esto dificulta extraer conclusiones, ya que algunos resultados inesperados podrían explicarse por el pequeño tamaño muestral. En relación al BAS-II infantil, la evidencia de validez convergente puede considerarse realmente escasa. Se muestran resultados favorables con las muestras originales, pero esto no puede considerarse suficiente, por lo que son necesarios más estudios.
2. La evidencia en relación a la validez de constructo factorial es generalmente positiva, pero los resultados son parcialmente contrarios en relación a Semejanzas gráficas. Puesto que los índices de aptitud intelectual se obtienen únicamente a partir de dos tests, la evidencia factorial debería ser más clara.
3. En relación al análisis de ítems, parece poco razonable que la mayoría de los ítems se ajusten al modelo de Rasch (lo que indicaría que todos los ítems son igualmente discriminativos). Esto puede explicarse porque las escalas ya han sido depuradas en sucesivas ediciones. Sin embargo, también podría deberse a que el tamaño de la muestra efectiva a la que se aplicó cada ítem no fuera suficiente para poder detectar esas diferencias. En el Manual falta información para valorar este punto. Resulta escasa la información que se ofrece sobre la comprobación de algunos supuestos necesarios a la hora de aplicar el modelo de TRI.
4. Podrían añadirse estudios de validez más específicos que muestren los perfiles de rendimiento de grupos con dificultades de aprendizaje específicas. No se incluyen estudios psicométricos en relación a la equidad de la prueba en relación a minorías étnicas en la muestra española. Tampoco se ofrece información sobre la fiabilidad test-retest para el BAS-II infantil en esta muestra.

En definitiva, la batería BAS-II es una excelente herramienta para obtener una descripción detallada y profunda de la inteligencia y del rendimiento cognitivo de niños y adolescentes.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test

BAS-II – Escalas de aptitudes intelectuales

1.2. Nombre del test en su versión original

British Ability Scales Second Edition (BAS-II)

1.3. Autor

Collin D. Elliott

1.4. Autor de la adaptación española

David Arribas Águila y Sara Corral Gregorio

1.5. Editor del test en su versión original

NferNelson Publishing Company Ltd. Londres

1.6. Editor de la adaptación española

TEA ediciones

1.7. Fecha de publicación del test original

1997.

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

2011

1.9. Fecha de la última revisión del test en su adaptación española

2011

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

- Inteligencia
- Aptitudes

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

El BAS-II es una batería para la evaluación del funcionamiento intelectual de los niños y los adolescentes. Los tests principales a aplicar varían en función de la edad y se organizan en dos baterías, BAS-II Infantil y BAS-II-escolar. Además de los tests principales, el BAS-II incluye tests diagnósticos, que evalúan aptitudes más específicas, y tests de rendimiento, relacionados con los logros educativos alcanzados. A partir de los tests principales permite obtener varios índices, que se describen a continuación

BAS-II Infantil
Nivel inferior, 2:6 a 3:5 años

Índice General (IG)	Resume la capacidad conceptual general y refleja la aptitud para realizar transformaciones complejas de la información (razonamiento y la aplicación de conceptos).
Índice General no Verbal (IGNV)	Se interpreta como IG, pero se calcula con los tests con menor componente verbal.

BAS-II Infantil
Nivel superior, 3:6 a 5:11 años

Índice General (IG)	Ver BAS-II infantil nivel inferior.
Índice General no Verbal (IGNV)	Ver BAS-II infantil nivel inferior.
Índice Verbal (IV)	Mide el conocimiento y la adquisición de conceptos verbales. Indica la capacidad de comprensión y expresión del lenguaje, el nivel de desarrollo del vocabulario y el conocimiento general básico.
Razonamiento Perceptivo (RP)	Implica la capacidad de procesar e integrar información visual y verbal, así como de recuperación del conocimiento previo almacenado en la memoria a largo plazo. También indica el conocimiento conceptual sobre los números, los hechos numéricos y las operaciones aritméticas simples.
Índice Espacial (IE)	Indica la capacidad del procesamiento visoespacial complejo (descomponer un diseño en sus componentes, reconstruir un todo a partir de sus partes componentes, percepción de la orientación espacial relativa, de la forma, la posición y el tamaño, etc.).

BAS-II Escolar: 6:0 a 17:11 años

Índice General (IG)	Ver BAS-II infantil nivel inferior.
Índice General no Verbal (IGNV)	Ver BAS-II infantil nivel inferior.
Índice Verbal (IV)	Ver BAS-II infantil nivel superior.
Razonamiento no verbal (RNV)	Implica la capacidad de razonamiento inductivo, es decir, la aptitud para identificar las reglas que gobiernan las características o las variaciones en las figuras abstractas o numéricas y la aptitud para formular y comprobar las hipótesis.
Índice Espacial (IE)	Ver BAS-II infantil nivel superior.

A continuación se describen los tests de cada batería (entre corchetes se muestran los índices a cuyo cálculo contribuyen):

BAS-II Infantil Nivel inferior, 2:6 a 3:5 años

Tests principales

Construcción	Habilidad visoperceptiva, especialmente orientación espacial, en la reproducción de modelos con bloques. [IG, IGNV]
Comprensión verbal	Lenguaje receptivo: comprensión de instrucciones orales relacionadas con conceptos lingüísticos básicos. [IG]
Semejanzas gráficas	Razonamiento no verbal en el emparejamiento de figuras que tienen en común un elemento o concepto. [IG, IGNV]
Nominación	Lenguaje expresivo; conocimiento de nombres. [IG]

BAS-II Infantil Nivel superior, 3:6 a 5:11 años

Tests principales

Comprensión verbal	ver BAS-II infantil nivel inferior. [IG, IV]
Semejanzas gráficas	ver BAS-II infantil nivel inferior. [IG, IGNV, RP]
Nominación	ver BAS-II infantil nivel inferior. [IG, IV]
Cubos	Razonamiento no verbal y visualización espacial en la reproducción de diseños con cubos de colores. [IG, IGNV, IE]
Conceptos numéricos	Conocimiento de conceptos prenuméricos y numéricos y resolución de problemas con ellos. [IG, RP]
Copia	Habilidad visoperceptiva y coordinación motora fina en la copia de dibujos lineales. [IG, IGNV, IE]

Tests diagnósticos

Construcción	ver BAS-II infantil nivel inferior.
Recuerdo de Objetos	Recuerdo visoespacial y verbal a corto y medio plazo de un conjunto de imágenes.
Dígitos de orden directo	Memoria auditiva a corto plazo y recuerdo oral de secuencias numéricas.
Emparejamiento de figuras	Discriminación visual de formas similares.
Reconocimiento de Imágenes	Memoria visual a corto plazo medida mediante el reconocimiento de objetos.
Dígitos orden inverso	Memoria de trabajo y recuerdo oral de secuencias numéricas inversas.

BAS-II Escolar: 6:0 a 17:11 años

Tests principales

Recuerdo de figuras	Recuerdo a corto plazo de relaciones visoespaciales mediante la reproducción de figuras abstractas. [IG, IGNV, IE]
Definiciones	Lenguaje expresivo: explicación del significado de palabras. [IG, IV]
Cubos	ver BAS-II infantil nivel superior. [IG, IGNV, IE]
Matrices	Razonamiento inductivo: identificación y aplicación de las reglas de relación que subyacen a formas abstractas. [IG, IGNV, RNV]
Semejanzas verbales	Razonamiento y conocimiento verbales. [IG, IV]
Razonamiento numérico	Razonamiento inductivo: detección y aplicación de reglas relativas a patrones secuenciales en fichas de dominó y relaciones entre pares de números. [IG, IGNV, RNV]

Tests diagnósticos

Recuerdo de Objetos	ver BAS-II infantil nivel superior.
Velocidad de procesamiento	Rapidez en la ejecución de operaciones mentales sencillas.
Dígitos de orden directo	ver BAS-II infantil nivel superior.
Reconocimiento de Imágenes	ver BAS-II infantil nivel superior.
Dígitos orden inverso	ver BAS-II infantil nivel superior.

Tests de rendimiento

Aritmética	Reconocimiento de números y ejecución de operaciones numéricas.
Ortografía	Manejo y conocimiento de las normas ortográficas básicas.

1.12. Áreas de aplicación

- Psicología clínica
- Psicología educativa
- Neuropsicología
- Servicios sociales
- Cualquier otra área en la que se requiera la evaluación de la inteligencia en niños y adolescentes

1.13. Formato de los ítems

- **Elección múltiple:** El evaluador presenta una lámina con el problema a resolver y el niño debe señalar la opción correcta: Semejanzas gráficas, Emparejamiento de figuras, Matrices.
- **Respuesta libre**, que puede ser oral (Nominación, Definiciones, Semejanzas verbales, Razonamiento numérico-juego B), escrita (Aritmética, Ortografía) o dibujada (Copia, Razonamiento numérico-juego A). Los ítems son presentados visualmente (Nominación, Razonamiento numérico, Copia, Aritmética) o verbalmente (Definiciones, Semejanzas verbales, Ortografía).
- **Presentación de modelos a imitar** (manipulativa): El evaluador muestra un modelo construido con bloques y el niño debe construir el mismo modelo: Construcción, Cubos.
- **Presentación visual de estímulos y tarea de recuerdo.** El evaluador presenta uno o más dibujos en una lámina. La tarea del niño puede ser nombrarlos (Recuerdo de objetos - ensayos verbales), disponerlos en la posición correcta (Recuerdo de objetos - ensayos espaciales), reconocerlos de entre un conjunto de distractores (Reconocimiento de imágenes) o dibujarlos (Recuerdo de figuras).
- **Presentación auditiva de estímulos y tarea de recuerdo.** El evaluador dicta una secuencia de dígitos. La tarea del niño es repetirla en el orden correcto: Dígitos (directo e inverso).
- **Acciones a partir de las instrucciones orales del aplicador.** El evaluador presenta láminas con dibujos y/o una serie de objetos. La tarea del niño es seguir las instrucciones del evaluador (señalar en una lámina, identificar/mover/organizar objetos, contar/enumerar fichas): Comprensión verbal, Conceptos numéricos.
- **Tachar la figura de cada fila con el mayor número de cuadrados o el número más alto:** Velocidad de procesamiento.

1.14. Número de ítems

El número de ítems es variable en función de la edad y el nivel de rendimiento que se va manifestando. A continuación, se señala el número de ítems disponible y, entre paréntesis, el mínimo y el máximo número de ítems puntuado, atendiendo a los bloques de puntuación preestablecidos.

Test	Ítems del test (mínima/máxima longitud)
Aritmética	46 (7/32)
Comprensión verbal	40 (12/31)
Conceptos numéricos	30 (11/27)
Construcción	16 (10/14)
Copia	20 (14/17)
Cubos	26 (10/20)
Definiciones	40 (14/30)
Dígitos de orden directo	36
Dígitos orden inverso	30
Emparejamiento de figuras	26 (18/26)
Matrices	33 (14/29)
Nominación	36 (17/31)
Ortografía	72 (26/61)
Razonamiento numérico	40 (14/27)
Reconocimiento de Imágenes	23 (12/18)
Recuerdo de figuras	16 (11/14)
Recuerdo de Objetos	2 ensayos, una lámina con 20 dibujos.
Recuerdo de Objetos	4 ensayos, una lámina con 20 dibujos.
Semejanzas gráficas	33 (22/33)
Semejanzas verbales	37 (13/24)
Velocidad de procesamiento	18 (6/6)

1.15. Soporte

- Administración oral
- Papel y lápiz
- Manipulativo

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

- Nivel C

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

- El BAS-II es aplicable a niños y adolescentes de edades comprendidas entre 2 años 6 meses y 17 años 11 meses. Se organiza en dos baterías: BAS-II infantil y BAS-II escolar. El BAS-II infantil permite evaluar a niños más pequeños (hasta los 5:11 años de edad) y el BAS-II escolar cubre las edades escolares (a partir de los 6 años). Por lo tanto, con ambas baterías se abarca un amplio rango de niveles educativos (educación infantil, primaria y secundaria). Excepcionalmente, pueden usarse tests del BAS-II infantil para niños mayores con dificultades o tests del BAS-escolar, para niños pequeños que muestran un rendimiento superior.
- El BAS-II es adecuado en cualquier situación en la que sea necesario evaluar la inteligencia y el funcionamiento cognitivo de niños o adolescentes. En general permite identificar los puntos fuertes y débiles del funcionamiento cognitivo, ya que sus tests se centran en los procesos cognitivos y poseen una alta especificidad. Parece adecuado tanto para identificar a niños con dificultades específicas de aprendizaje como para el diagnóstico de retraso mental y altas capacidades.

- Para niños con dificultades de aprendizaje severas o profundas, se puede adaptar la dificultad y aplicar pruebas fuera de rango. En la mayoría de las ocasiones existen baremos para su interpretación, aunque también existe la posibilidad de utilizar las tablas de edades equivalentes, que tienen conocidas limitaciones, por lo que deben considerarse sólo a nivel orientativo.
- Si se sospecha o conoce algún déficit verbal o auditivo (p.ej., pérdida auditiva severa, problemas del habla y del lenguaje, niños procedentes de ambientes familiares carentes de estimulación verbal o de habla extranjera, niños extremadamente tímidos) pueden aplicarse las pruebas no verbales y utilizar un índice general no verbal (IGNV). El evaluador puede transmitir las instrucciones mediante gestos y el niño puede responder de forma no verbal, mediante señales o manipulando objetos. Los tests principales no verbales son Construcción, Semejanzas gráficas, Cubos, Copia, Matrices, Razonamiento numérico y Recuerdo de figuras.
- En principio, puede ser aplicado a minorías étnicas ya que la batería no incluye en general contenidos específicos de un país o cuestiones sobre valores sociales. Sin embargo, los tests verbales (Nominación, Dígitos, Recuerdo de objetos, Comprensión verbal, Semejanzas verbales, Definiciones y Ortografía) no deben ser aplicados (ni traducidos) a niños poco competentes en el idioma español (p. 66 del Manual de aplicación). Para estos casos, el BAS-II incorpora la posibilidad de realizar una aplicación de tipo no verbal, incluyendo los tests que no se ven afectados por esta característica y permitiendo el cálculo del Índice General no Verbal (IGNV). El Manual técnico (p. 297) recomienda que se sea cauto en la interpretación de estas puntuaciones en niños recién llegados a la cultura de destino, si proceden de entornos donde no sea habitual dibujar, copiar patrones, manipular cubos o trabajar con límite de tiempo.
- Algunas pruebas pueden subestimar el funcionamiento cognitivo en el caso de déficits motores (p.ej., Copia, Cubos, Recuerdo de figuras, Razonamiento numérico –conjunto A), falta de atención o distractibilidad (p.ej., Construcción, Comprensión verbal, Recuerdo de figuras, Reconocimiento de imágenes, Velocidad de procesamiento, Aritmética), ansiedad al ser cronometrado (p.ej., Cubos-procedimiento estándar), problemas perceptivos (p.ej., Copia) Impulsividad (p.ej., Comprensión verbal, Semejanzas Gráficas, Matrices, Razonamiento numérico, Emparejamiento de figuras) o un balance inadecuado de velocidad y precisión (velocidad de procesamiento). El manual es rico en detalles de estas y otras explicaciones alternativas para explicar el rendimiento bajo en las tareas (pp. 80-135 del Manual técnico).

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

- No existen formas paralelas ni tampoco versiones abreviadas específicas, aunque el test contempla una aplicación de los tests principales (4 en el BAS-II infantil-nivel inferior, 6 en el BAS-II infantil-nivel superior, 6 en el BAS II-escolar) o de la batería completa (6 en el BAS-II infantil-nivel inferior, 12 en el BAS-II infantil-nivel superior, 11 en el BAS II-escolar).

1.19. Procedimiento de corrección

- Manual mediante plantilla
- corrección online y los descritos en la siguiente nota

NOTA: En la mayoría de las pruebas se puntúa como acierto o error según la plantilla o los criterios que aparecen en el cuadernillo de anotación y en el Manual de aplicación. La corrección debe hacerse en el momento, para poder evaluar si la aplicación se detiene o continúa. En general, la aplicación y corrección de los ítems es sencilla. Para algunas pruebas se incluye la solución correcta en el propio cuadernillo de anotación (Semejanzas gráficas, Conceptos numéricos, Emparejamiento de figuras, Reconocimiento de imágenes, Matrices, Razonamiento numérico, Aritmética, Ortografía). En otras pruebas, el evaluador debe valorar si un modelo construido es correcto (Construcción, Cubos), si se ha seguido una instrucción verbal correctamente (Comprensión verbal), si la secuencia de dígitos se ha recordado correctamente (Dígitos directo y Dígitos inverso), si ha marcado el elemento numérico más alto (Velocidad de procesamiento) o anotar el número de elementos recordados o correctamente dispuestos en la lámina (Recuerdo de objetos). En algunas pruebas se requiere el cómputo del tiempo de respuesta para establecer la puntuación (Cubos, Velocidad de procesamiento), pero en ninguno de estos casos se requiere una valoración compleja de la respuesta.

Existen algunas excepciones en las que la corrección es más compleja, ya sea porque ciertas respuestas, definidas en el manual de aplicación, requieren preguntas de aclaración por parte del evaluador (Nominación, Definiciones y Semejanzas Verbales) o porque los criterios de corrección son muy detallados (p.ej., puntuación no dicotómica) y requieren un cierto entrenamiento previo (Copia y Recuerdo de figuras).

La editorial ha desarrollado un sistema online (www.teacorrige.com) que permite la corrección automática.

1.20. Procedimiento de obtención de las puntuaciones directas

- En general se obtienen las puntuaciones directas (PD) sumando los aciertos a los ítems aplicados. Siguen este modo de puntuación Construcción, Comprensión verbal, Semejanzas gráficas, Nominación, Conceptos numéricos, Emparejamiento de figuras, Reconocimiento de imágenes, Definiciones, Semejanzas verbales, Matrices, Razonamiento numérico, Aritmética y Ortografía.

- Se dan las siguientes excepciones o variaciones:

En Velocidad de procesamiento y en Cubos, se puntúa positivamente la velocidad de la respuesta. En Velocidad de procesamiento se puntúa en 7 niveles (de 0 a 6, según el tiempo de respuesta). En Cubos (puntuación estándar) se pondera, dependiendo del ítem, en cinco (de 0 a 4) o seis niveles (de 0 a 5).

En Copia y Recuerdo de figuras se pondera en función de la precisión o calidad de la respuesta. Se puntúa en tres (0, 1, 2) o cuatro categorías (0, 1, 2, 3) según el ítem.

En Recuerdo de objetos inmediato, se puntúa el número de objetos de una lámina correctamente recordados en un tiempo limitado. Se realizan tres ensayos con la misma lámina. En un cuarto ensayo, se puntúa que los evaluados recuerden correctamente la posición del objeto en la lámina.

En Recuerdo de objetos demorado se procede como en Recuerdo de objetos inmediato pero sin presentación previa.

En Dígitos de orden directo y Dígitos de orden inverso, el niño también recibe puntos por los ítems anteriores al bloque base, aunque no se le hayan aplicado.

Las puntuaciones directas (PD) obtenidas no son interpretables pues para diferentes niños pueden haber sido obtenidas con distintos bloques de ítems (i.e., los ítems están ordenados por dificultad y, según la edad y el rendimiento que va manifestando el niño, varían el punto de comienzo y el punto de terminación del test).

Las PD se transforman a Puntuaciones de Aptitud (PA), que se obtienen a partir de un modelo de Teoría de la Respuesta al Ítem (en concreto, el modelo de Rasch). Este proceso se describe en el Manual técnico (p. 324).

El Manual proporciona para cada test una tabla de correspondencia entre PD y PA según el bloque de ítems aplicados, por lo que para la puntuación de los tests no es necesario ni un programa informático ni un conocimiento profundo de los modelos de TRI. En todo caso, se puede hacer uso del sistema de corrección automática, online, que el editor ofrece.

1.21. Transformación de las puntuaciones

- Normalizada

1.22. Escalas utilizadas

- Centiles
- T (*Media 50 y desviación típica 10*)
- Otra (Véase NOTA)

- **NOTA:** Se utilizan además la escala CI (*Media 100 y desviación típica 15*) y las edades equivalentes.

La puntuación de aptitud (PA) no es una puntuación normativa ya que el nivel de cada evaluado debe referenciarse a su grupo de edad. A partir de la PA se obtienen para cada test cognitivo las

puntuaciones T y los percentiles tomando como referencia la distribución de PA en los niños de la misma edad.

La escala de “edades equivalentes” indica la edad en la que la puntuación de aptitud del niño equivale a la mediana de la distribución. El Manual señala las importantes limitaciones de esa escala y aconseja precaución en su uso.

Las puntuaciones en los índices se obtienen sumando las puntuaciones T de los tests correspondientes, para posteriormente obtener los centiles y las puntuaciones de CI asociadas.

En todos los casos, los centiles, las puntuaciones T y las puntuaciones de CI se obtienen mediante un procedimiento denominado “*continuous norming*” (ver apartado Baremación, en la página 23).

1.23. Posibilidad de obtener informes automatizados

- No

1.24. Oferta por parte del editor de un servicio para la corrección y/o elaboración informes

- Si

La editorial ha desarrollado un sistema online que permite obtener automáticamente el perfil de resultados, las puntuaciones CI, T y percentiles, así como el cálculo de las discrepancias, su significación y frecuencias.

1.25. Tiempo estimado para la aplicación del test

- La aplicación es individual. La aplicación de los tests principales requiere entre 25 y 45 minutos. La aplicación de la batería completa (los tests principales más los diagnósticos, entre 35 y 70 minutos).

1.26. Documentación aportada por el editor

- Manual
- Otra (Véase NOTA)

NOTA: En la página web de TEA ediciones (<http://web.teaediciones.com>) se anuncia un libro en preparación sobre la evaluación con el BAS-II:

Calonge, I. y Arribas, D. (en preparación). *Claves para la evaluación con el BAS-II*. Madrid: TEA ediciones.

1.27. Precio de un juego completo de la prueba

- Juego Completo BAS-II Infantil (713,96 €)
- Juego Completo BAS-II Escolar (632,84 €)
- Juego Completo BAS-II Infantil+Escolar (977,60 €)

1.28. Precio y número de ejemplares del paquete de cuadernillos

- Ejemplar Razonamiento numérico (RN) A Escolar (104,00 €)
- Ejemplar Velocidad de procesamiento (VP) A Escolar (31,20 €)
- Ejemplar Velocidad de procesamiento (VP) B Escolar (31,20 €)
- Ejemplar Velocidad de procesamiento (VP) C Escolar (31,20 €)
- Ejemplar Aritmética y ortografía (A/O) Escolar (46,80 €)
- En todos los casos el precio es por paquete de 25 ejemplares.

1.29. Precio y número de ejemplares del paquete de hojas de respuesta

- Cuadernillo de anotación INFANTIL (72,80 €)
- Cuadernillo de anotación ESCOLAR (RF) (72,80 €)
- En los dos casos el precio es por paquete de 25 ejemplares.

1.30. Precio de la corrección y/o elaboración de informes por parte del editor

La corrección online es gratuita para los usuarios del BAS-II.

1.31. Bibliografía básica acerca del test aportada en la documentación

La bibliografía es extensa pero no hay recomendaciones sobre cuáles son las referencias básicas y específicas para esta prueba. Por nuestra parte, una referencia básica a considerar (no mencionada en el Manual) puede ser Elliott (2001).

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

Aspectos generales

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Buena	4
2.2. Calidad de la documentación aportada	Excelente	5
2.3. Fundamentación teórica	Excelente	5
2.4. Adaptación del test	Excelente	5
2.5. Calidad de las instrucciones	Excelente	5
2.6. Facilidad para comprender la tarea	Excelente	5
2.7. Facilidad para registrar las respuestas	Excelente	5
2.8. Calidad de los ítems (aspectos formales)	Excelente	5
2.9. Datos sobre el análisis de los ítems	Excelente	5

En general la edición y manufactura de los materiales es muy buena, aunque se han detectado algunas imperfecciones de muy escasa importancia. Las pestañas y portadas de los libros de láminas se doblan y rompen con cierta facilidad. La mochila resulta algo grande y vendría bien que fuese más rígida para asegurar una mejor protección de su contenido. En relación al Manual técnico hay una errata en la fórmula de la página 220 y una tipografía excesivamente reducida para los rótulos del eje de ordenadas de las figuras 8.2 a 8.4 (funciones de información), que dificulta su lectura. En relación al juego de estímulos, existen algunas pequeñas imperfecciones en Cubos (en algunos cubos, el color de una cara lateral tiene cierto relieve por lo que se forma una estrecha franja de ese color en la cara superior), Conceptos numéricos y Comprensión verbal (algunas fichas no son completamente lisas) y Recuerdo de objetos (las tarjetas están ligeramente teñidas de un tono azul). No se espera que las imperfecciones señaladas afecten al rendimiento de los evaluados en el test.

Dada la complejidad del diseño (i.e., la aplicación adaptativa), el análisis de ítems resulta difícil de valorar desde las estrategias de análisis clásico. Se realizó un estudio piloto (N = 552) con algunas de las pruebas obteniendo los índices de dificultad y discriminación clásicos y, mediante el modelo psicométrico de Rasch, el parámetro *b* de dificultad y un indicador de ajuste para cada ítem. Los resultados (no detallados en el Manual) llevan a los autores a rechazar una de las pruebas que no consta en la versión definitiva (Lectura de palabras). También se eliminaron

algunos ítems de Definiciones por su falta de ajuste al modelo de Rasch, se ordenaron los ítems según el parámetro b obtenido y se modificaron parcialmente los criterios de inicio y parada.

En la muestra de baremación ($N = 1725$) se aplicó el modelo de Rasch y se comprobó el ajuste de los ítems al modelo. No parece que se eliminen ítems en esta fase, aunque sí se reordenan, en general, por su dificultad y se elaboran los puntos de inicio y decisión definitivos. Una limitación de estos análisis es que no se concreta cuál es la muestra efectiva a la que se aplicó cada ítem o la longitud promedio de los tests en que ese ítem fue aplicado.

Validez de contenido

Contenido	Valoración	Puntuación
2.10. Calidad de la representación del contenido o dominio	Buena	4
2.11. Consultas a expertos	Se ha consultado a un número moderado de expertos mediante un procedimiento sistematizado ($10 \leq N \leq 30$)	4

Validez de constructo

Contenido	Valoración	Puntuación
2.12. Diseños empleados	<ul style="list-style-type: none"> ▪ Correlaciones con otros tests ▪ Análisis factorial exploratorio ▪ Análisis factorial confirmatorio ▪ Estudio de validez cruzada del IG a través de las edades (este estudio se realiza con una batería similar al BAS-II, el DAS) 	
2.13. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado	4
2.14. Procedimiento de selección de las muestras	Véase NOTA	
2.15. Correlaciones del test con otros tests similares	Excelente ($r \geq 0.60$)	5
2.16. Calidad de los test empleados como criterio o marcador	Excelente	5
2.17. Datos sobre el sesgo de los ítems	Excelente	5

NOTA: La tipificación española del BAS-II siguió las especificaciones y proporciones del censo de población española de 2008. Se usaron 4 criterios para diseñar el plan de muestreo: la edad (16 grupos de edad entre 2:6 y 17:11), el sexo, la región geográfica (las autonomías se agruparon en 4 regiones: Norte, Sur, Este y Centro) y el país de procedencia (se calculó el porcentaje de niños extranjeros según los datos de escolarización del censo y fue común para todas las edades: 7,6% de la muestra). El plan diseñado contempló recoger 1700 casos. La muestra a recoger se distribuyó entre los colaboradores de cada región. Además se recogió información sobre: la zona de residencia (urbana, intermedia, rural); tipo de centro escolar al que asistía el niño (público, concertado o privado); nivel socioeconómico y educativo de los padres; lengua materna (castellano, catalán, vascuence o gallego); nacionalidad y dos variables de control (presencia de problemas sensoriales, de lenguaje o motores y presencia de dificultades de aprendizaje).

Validez predictiva

- En relación a la muestra española, no hay un apartado de validez predictiva como tal en el Manual técnico. Sin embargo, pueden considerarse como criterios de rendimiento las pruebas estandarizadas de Ortografía y Aritmética, que fueron aplicadas a la muestra de baremación del BAS-II escolar.

Contenido	Contenido	Puntuación
2.18. Diseño de selección del criterio	<ul style="list-style-type: none"> Concurrente 	
2.19. Tamaño de las muestras	Varios estudios con muestras representativas de tamaño moderado	4
2.20. Procedimiento de selección de las muestras		-
2.21. Mediana de las correlaciones del test con los criterios	Muy buena ($0.45 \leq r < 0.55$)	4

Comentarios generales sobre validez

Validez de contenido: Cabe destacar que la versión original del BAS-II se basa en una batería previa (el DAS), pero introduce diversas modificaciones que facilitan la aplicación o mejoran la interpretación de las puntuaciones. En la edición española, se percibe un esfuerzo importante (estudio piloto, consultas a expertos) por reemplazar ítems o modificar su posición, atendiendo a nuestras características culturales e idiomáticas. El material detalla adecuadamente las

funciones y procesos implicados por cada test y la composición de la batería se sustenta en el modelo jerárquico de la inteligencia de Cattell-Horn y Carroll. En su conjunto, los tests representan razonablemente el dominio del modelo citado. Una posible limitación es que los índices de primer orden (IGV, RP, RNV e IE) se obtienen únicamente a partir de dos tareas.

Validez de constructo factorial: Se realiza un análisis factorial confirmatorio (AFC) para cada grupo de edad y los resultados permiten defender razonablemente la estructura teórica propuesta en relación a los tests principales e índices globales. Cabe destacar como limitación menor que los índices de ajuste elegidos no son los óptimos; hubiera sido preferible el uso del TLI y el RMSEA, frente al GFI y al RMSR. En relación a los análisis factoriales exploratorios (AFE), en el que se analizan los tests principales y algunos tests diagnósticos, los resultados son generalmente satisfactorios, aunque un resultado es incongruente con las expectativas teóricas: en la batería BAS II infantil-nivel superior, Semejanzas gráficas no pesa en el factor de razonamiento perceptivo (peso = 0.02) y tiene mayor peso en los factores visoespacial (peso = 0.32) y verbal (peso = 0.34). Este resultado evidencia que son necesarios más estudios en relación a la estructura multidimensional del BAS-infantil superior en relación a este test. Como limitaciones generales de los análisis AFE cabe señalar que: a) no se hace uso de las técnicas disponibles para decidir el número óptimo de factores (p.ej. el análisis paralelo), por lo que es difícil discriminar cuál es la estructura más adecuada; y b) no está claro por qué los tamaños muestrales son superiores en el AFE, lo que quizás se deba a un diferente tratamiento de los valores perdidos que no se explicita.

Relaciones con otros tests: En la muestra de baremación española se realizan distintos estudios de validez concurrente con pequeñas muestras de distintas edades. En general los resultados fueron satisfactorios.

En relación al WISC-IV se realizó un estudio con el BAS-II escolar (N = 24) se encontraron excelentes correlaciones de IG y IGV con el CI total (0.84 y 0.76, respectivamente), de IV con Comprensión verbal (0.85), de IE con Razonamiento perceptivo y Memoria de trabajo (0.71, 0.74) y de RNV con Comprensión verbal y Velocidad de procesamiento (0.71, 0.73). Sin embargo, también se encontraron algunos resultados inesperados que quizás se deban al pequeño tamaño muestral del estudio. También se encontraron correlaciones muy altas (superiores a 0.79) entre los índices y la puntuación total del RIST (Test de Inteligencia Breve de Reynolds).

En relación al BAS-II infantil sólo se incluye un estudio con la muestra española, pero la muestra es excesivamente pequeña (N = 13) para poder extraer ninguna conclusión.

Se incluyen estudios, realizados con el test original, del BAS-II infantil con el WPPSI-R (N = 44) y con el BAS-R infantil (N = 48), y del BAS-II escolar con el WISC-III (N = 38) y con el BAS-R escolar (N = 31). En general, los índices y los tests del BAS-II correlacionaban satisfactoriamente y de la forma esperada con los índices de las escalas Wechsler. Sólo dos tests diagnósticos, Recuerdo de

objetos y Recuerdo de imágenes mostraron correlaciones menores en el grupo de menor edad (correlaciones entre 0.12 y 0.41).

Finalmente, también la muestra original, se realiza un análisis de la equivalencia entre los IG obtenidos en el BAS-II infantil y el BAS-II escolar para una muestra de 400 niños con un rango de edad en la que ambas baterías son aplicables. Este análisis se realiza con la versión de los tests aplicada en el DAS (que es similar a la del BAS-II). Los resultados de un AFC muestran que puede asumirse que ambos IG son equivalentes.

Estudios del sesgo: Se dedica un capítulo completo al estudio de la equidad del BAS-II. Se realizó la revisión por expertos del contenido. También se describen estudios psicométricos específicos de detección del DIF en relación a la etnia (en la muestra original) y con respecto al sexo (en las muestras original y española). Finalmente, se incluye un estudio de la invarianza de la validez predictiva (a través de la etnia) en la muestra original utilizando como criterio una escala tipificada de comprensión lectora de frases (Suffolk Reading Scale).

Validez predictiva: En relación a la muestra española, no hay un apartado de validez predictiva como tal en el Manual técnico. Sin embargo, pueden considerarse como criterios de rendimiento las pruebas estandarizadas de Ortografía y Aritmética, que fueron aplicadas a la muestra de baremación del BAS-II escolar. La correlación entre ambos criterios es de 0.42 y su consistencia interna alta (superior a 0.80). Para el criterio Aritmética se encuentran correlaciones excelentes con IG, IGNV y con RNV (0.55, 0.52 y 0.53) y entre los tests el mejor predictor es Razonamiento numérico ($r = 0.50$). Para el criterio Ortografía se encuentran correlaciones buenas con IG, IGNV, IV y con RNV (0.42, 0.39, 0.35 y 0.37).

Fiabilidad

Contenido	Valoración	Puntuación
2.22. Datos aportados sobre fiabilidad	<ul style="list-style-type: none"> ▪ Coeficientes de fiabilidad para diferentes grupos de sujetos ▪ Error típico de medida para diferentes grupos de sujetos 	

Equivalencia formas paralelas

Contenido	Valoración	Puntuación
2.23. Tamaño de las muestras	No se aporta	-
2.24. Coeficientes de equivalencia	No se aporta	-

Consistencia interna

Contenido	Valoración	Puntuación
2.25. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado	4
2.26. Coeficientes de consistencia	<ul style="list-style-type: none"> ▪ Buena ($0.80 \leq r < 0.85$) (Para los tests) ▪ Excelente ($r \geq 0.85$) (Para los índices) 	

La muestra total es de aproximadamente 1700 casos, pero los coeficientes se calculan para 18 grupos de edad (en general, de aproximadamente 100 personas).

Estabilidad (test-retest)

Contenido	Valoración	Puntuación
2.27. Tamaño de las muestras	Un estudio con una muestra pequeña ($N < 100$)	1
2.28. Mediana de los Coeficientes de estabilidad	Excelente ($r \geq 0.80$) Para tests e índices en el BAS-escolar	5

- Para el BAS-infantil no se dispone de resultados en la muestra española.

Comentarios generales sobre fiabilidad

- Para los índices, los indicadores de consistencia interna son excelentes en casi todos los casos. La única excepción ocurre en el grupo de menor edad (2:6 a 3:5 años), en el que la consistencia interna del índice IGNV puede etiquetarse como adecuada. Esta excepción se señala explícitamente en el manual, que recomienda el uso de IGNV sólo en aquellos casos en los que IG resulte inadecuado por su contenido verbal.
- En general la consistencia interna de los tests individuales es buena y en algunos casos (p.ej., Construcción, Cubos, Dígitos orden directo, Dígitos orden inverso, Conceptos numéricos, Aritmética y Ortografía), excelente. Algunas excepciones a esta tendencia se encuentran para Nominación, Semejanzas gráficas, Semejanzas verbales, Razonamiento numérico y Recuerdo de Figuras, en los que la consistencia interna puede calificarse sólo de

adecuada en más de la mitad de los grupos de edad analizados (entre 0.70 y 0.79). El peor resultado se obtiene para Reconocimiento de imágenes (menor de 0.7 para diversos grupos de edad) y, para sujetos con edades menores de 5 años, en la prueba Recuerdo de objetos inmediato, donde el coeficiente alfa se encuentra entre 0.57 y 0.65. En dos tests se calculó el coeficiente alfa. En el resto, se calculó el coeficiente de fiabilidad por el método de las dos mitades, tras aplicar un ingenioso y psicométricamente avanzado procedimiento.

- Otro punto valorable es que se informa de la función de información de los tests, especificando el rango de nivel de rasgo en el que se aplica el test habitualmente. Esto permite valorar la precisión de la prueba para cada nivel de rasgo.
- Se informa de la estabilidad para una muestra de 56 niños entre 6:2 y 12:11 años de edad, con un intervalo de entre 2 y 8 semanas (en promedio, un mes), utilizando el mismo evaluador en la mayoría de los casos. Los resultados son excelentes para los índices y para los tests, con alguna excepción, como Recuerdo de Figuras (0.62). Como es lógico, se observan efectos de la práctica en varios tests, por lo que los autores recomiendan tenerlos en cuenta cuando se realiza una reevaluación después de unos pocos meses de la evaluación inicial (pp. 238-239 del Manual técnico).
- Para el BAS infantil no se dispone de información en la muestra española. Se informa de los resultados para una versión previa (el DAS) en una muestra distinta de la española. Los resultados (para dos muestras de aproximadamente 100 sujetos, con rangos de edad 3:6-4:5 y 5:0-6:3 respectivamente) son muy variados.
- Conviene destacar que se tenga en cuenta la fiabilidad interjueces para evaluar la calidad de la corrección en las pruebas de respuesta abierta (Copia, Recuerdo de figuras, Semejanzas y Definiciones), si bien los autores de la adaptación remiten al estudio original con la versión previa (el DAS) para valorar la fiabilidad interjueces. En ese estudio previo, con 4 muestras pequeñas para cada test, encuentran que la fiabilidad es generalmente excelente (mediana de la correlación intraclase para el acuerdo absoluto, 0.95).
- Para el BAS-II, sólo se realizó un análisis del test Recuerdo de figuras (dado que para esta prueba se habían dado los mayores cambios en los criterios de corrección). Los autores encuentran que, con los criterios finales de corrección, el porcentaje de acuerdo entre las evaluaciones dadas por los correctores y por un corrector experto, para una muestra de 50 casos, fue del 93%.

Baremación

Contenido	Valoración	Puntuación
2.29. Calidad de las normas	Amplio rango de baremos en función de la edad, el sexo, el nivel cultural y otras características relevantes	5
2.30. Tamaño de las muestras	Moderado (300 ≤ N < 600)	3
2.31. Procedimiento de selección de las muestras	Véase nota	

- **NOTA:** Se atendió a cuatro criterios a la hora de diseñar el plan de muestreo: la edad (16 grupos de edad entre los 2:6 y los 17:11 años), el sexo (50% de cada sexo), la región geográfica (Norte, Sur, Este y Centro, respetando las proporciones del censo de población) y el país de procedencia (7.6% de nacionalidad extranjera, teniendo en cuenta los datos de escolarización del censo). La obtención de muestra se llevó a cabo a través de 53 colaboradores, que realizaron entre 40 y 56 aplicaciones. Se recogió información adicional: zona de residencia del niño (urbana, intermedia, rural), tipo de centro escolar al que asistía (público concertado, privado), nivel socio-económico y educativo de los padres, lengua materna (Castellano, catalán, vascuence o gallego), nacionalidad y dos informaciones de control: presencia de problemas auditivos, de visión, lenguaje o motor y dificultades de aprendizaje (dislexia, discalculia...). Las muestras recogidas finales permiten concluir que la muestra es razonablemente representativa del censo de población (2008) en relación a las características señaladas.

Comentarios generales sobre baremación

Se ofrecen baremos distintos para cada grupo de edad (de aproximadamente 150 niños). Cabe destacar que se utiliza un procedimiento óptimo de baremación ("continuous norming") que establece un modelo para aproximar la función de crecimiento de cada percentil a través de la edad. De esta manera, se reduce el error asociado, pues en el cálculo del percentil para un grupo de edad se considera lo obtenido en los otros grupos de edad. Una limitación es que no se informa de cuál sería el tamaño "equivalente" en relación a un procedimiento tradicional. En cualquier caso, puede entenderse que el procedimiento utilizado podría ser casi equivalente a lo obtenido triplicando las muestras en el uso de un procedimiento de baremación tradicional (Evers, Sijtsma, Lucassen & Meijer, 2010).

Referencias

Elliott, C.D. (2001). *Application of the Differential Ability Scales (DAS) and British Ability Scales Second Edition (BAS-II), for the assessment of learning difficulties*. En A.S. Kauffman y N. Kauffman, *Specific learning disabilities and difficulties in children and adolescents*. Cambridge, Cambridge University Press.

Evers, A., Sijtsma, K, Lucassen, W. and Meijer, R. (2010). The Dutch Review Process for Evaluating the Quality of Psychological Tests: History, Procedure, and Results, *International Journal of Testing*, 10: 4, 295-317.