

# Evaluación de

## Batería de Actividades mentales Diferenciales y Generales, Nivel i

 **BADyG/i** 



# RESUMEN DE LA VALORACIÓN DEL TEST

## Descripción general

Característica	Descripción
Nombre del test	BADyG/i: Batería de Actividades mentales Diferenciales y Generales, Nivel infantil
Autor	Carlos Yuste Hernanz y David Yuste Peña
Autor de la adaptación española	-
Variable(s)	Inteligencia, habilidades
Áreas de aplicación	Psicología educativa
Soporte	Papel y lápiz

## Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Buena-Excelente	4,5
Fundamentación teórica	Excelente	5
Adaptación	-	-
Análisis de ítems	Bueno	4
Validez: contenido	Buena	4
Validez: relación con otras variables	Adecuada-Buena	3,5
Validez: estructura interna	Adecuada, pero con algunas carencias	2,5
Validez: análisis del DIF	-	-
Fiabilidad: equivalencia	-	-
Fiabilidad: consistencia interna	Buena	4
Fiabilidad: estabilidad	-	-
Fiabilidad: TRI	-	-
Fiabilidad inter-jueces	-	-
Baremos e interpretación de puntuaciones	Buena - Excelente	4,7

Nota. El signo – se interpreta como que no se aporta información o bien que no procede.

## Comentarios generales

El BADyG/i es una prueba de rendimiento óptimo de aplicación individual o semicolectiva destinada a evaluar inteligencia en escolares de entre 4 y 6 años de edad y comprende un conjunto de 6 pruebas básicas y 2 pruebas complementarias. Basado en una concepción jerárquica de la inteligencia, las 6 pruebas de aptitudes básicas se agrupan en dos factores generales de inteligencia, verbal (IGV) y no verbal (IGNV), que a su vez conforman un factor de orden superior de inteligencia general (IG). La principal área de aplicación es la orientación escolar en el contexto educativo.

Dispone de un manual técnico impreso y un extenso manual teórico-técnico completo (MTTC) en formato digital accesible desde la plataforma on-line Gestor BADyG. Las instrucciones para la administración de la prueba en soporte lápiz y papel son muy detalladas y claras. Para facilitar la tarea a quien responde el test, se dispone de un cuaderno de elementos-respuestas, en el que simplemente se debe marcar la opción seleccionada con una cruz. En la documentación también se ofrece la información completa para la corrección de cada prueba, que puede hacerse de forma manual o automatizada previa introducción de las respuestas en el gestor on-line. Además, con esta última opción, el gestor proporciona hasta seis tipos de informes, tanto individuales como colectivos, que resultan muy completos, informativos y útiles para detectar los puntos fuertes y débiles del desarrollo cognitivo, y en caso de limitaciones en el aprendizaje poder orientar una consecuente intervención.

El MTTC ofrece una extensa y sólida justificación teórica del constructo evaluado y sus facetas. También se proporcionan análisis de tareas que ilustran el proceso de respuesta previsto en cada prueba. En cuanto a las evidencias de validez basadas en el contenido, son de tipo teórico y la depuración de la versión experimental para llegar a la versión actual se reduce a un análisis cuantitativo de los elementos. En este análisis de ítems obtenido con la actualización de las muestras normativas se observa que hay diversos ítems que no alcanzan los criterios fijados por los autores (mínimo ,30 de homogeneidad y 2% de frecuencia de respuesta de cada alternativa). Convendría dejar explícito en el manual las estrategias y procedimientos seguidos para garantizar que se han incluido los ítems pertinentes en cuanto a relevancia y representatividad. Dado que en el manual impreso se explicita que se van a renovar las pruebas, sería aconsejable que al hacerlo se detalle más el proceso de generación de los elementos y se lleve a cabo también un análisis cualitativo de los mismos, teniendo en cuenta otros aspectos, principalmente la aportación de evidencias basadas en el juicio de especialistas en los constructos evaluados además de los propios autores, y adicionalmente la realización de una prueba piloto para observar la reacción del alumnado a quién va dirigido el test y la opinión del profesorado.

Las evidencias de validez basadas en la estructura interna se centran en la dimensionalidad de las 6 pruebas básicas, y el análisis en componentes principales realizado en una muestra muy grande da soporte empírico al modelo jerárquico hipotetizado de forma aceptable, aunque la información proporcionada sería susceptible de ser ampliada. Sería interesante replicar el modelo obtenido con AFE (análisis factorial exploratorio) o ESEM, para así disponer de índices de ajuste más potentes ( $\chi^2$ -cuadrado y derivados).

Respecto a las evidencias basadas en la relación con variables externas, únicamente se ofrece información sobre la relación de las puntuaciones con la variable sexo en una muestra grande. Se echa en falta el estudio de la relación con la edad, para aportar evidencias que las puntuaciones difieren de 6 en 6 meses. También se sugiere incorporar al nivel infantil evidencias en relación al criterio externo de rendimiento académico contemplado en otros niveles de las baterías BADyG, ya sea con una medida cuantitativa o con una categórica con unos pocos niveles a partir del juicio subjetivo del profesorado. Otra posible mejora podría ir orientada a proporcionar evidencias de validez convergente y discriminante con otro instrumento disponible, que podría aplicarse a una submuestra.

La fiabilidad se analiza desde la Teoría Clásica de los Tests y se proporcionan diversos indicadores de consistencia interna con muestras grandes segmentadas por edad. Por un lado, los valores de alfa de Cronbach son entre excelentes y buenos para las puntuaciones de Inteligencia General y los factores de Inteligencia Verbal y no Verbal (entre ,82 y ,92) y más discretos (entre ,55 y ,78) para las 6 pruebas de aptitudes básicas, lógicamente de menor longitud (18 ítems). El valor

promedio de alfa para el conjunto de todas las medidas es de ,74. No obstante, convendría discutir el uso del coeficiente alfa de Cronbach en un contexto donde los ítems de las pruebas se presentan ordenados por dificultad. Por otro lado, también se utiliza la estrategia de dos mitades para las puntuaciones de los factores de segundo y tercer orden, que podría constituir una aproximación a la equivalencia de medidas paralelas si ambas mitades se consideran como dos medidas completas pero reducidas. Por ello, sería aconsejable aportar también este indicador para las escalas de primer orden. Como aspecto positivo cabe destacar que también se ofrecen los valores del error máximo de todas las medidas (excepto de las 2 pruebas complementarias), que permiten estimar los intervalos de confianza de las puntuaciones individuales obtenidas. Sería conveniente aportar información del acuerdo entre evaluaciones en el caso de las 2 pruebas complementarias de respuesta construida, para proporcionar evidencias basadas en la fiabilidad inter-jueces, al menos en una submuestra. Adicionalmente, se podría valorar la conveniencia de emplear otras estrategias de evaluación de la fiabilidad de las puntuaciones, como por ejemplo mediante el diseño de un estudio de estabilidad temporal, para proporcionar evidencias basadas en la fiabilidad test-retest en un intervalo breve entre administraciones, pudiéndose llevar a cabo en una parte de la muestra.

Se proporcionan 6 baremos mixtos de 6 en 6 meses, aunque en el manual se especifica que se obtienen en función del grupo o curso y no en función de la edad cronológica, con transformaciones a percentiles y eneatis; también es posible obtener el CI (Cociente Intelectual) de la puntuación total (IG) a partir de la edad cronológica. Es destacable que las muestras utilizadas para la baremación son entre muy grandes y grandes (de más de 600 participantes para cada grupo de edad considerado) y actuales (última actualización de 2016). Incluyen alumnado de entre 4 y 6 años de edad de escuelas públicas, concertadas y privadas y de diversas comunidades autónomas, abarcando el rango de edad al que va destinado el test. Sería deseable incorporar en el manual impreso los resultados de las comparaciones por sexo que ahora sólo se encuentran en formato digital, añadiendo además el cálculo del tamaño del efecto de las diferencias para todas las medidas, para visibilizar mejor que no existen diferencias relevantes, así como proporcionar resultados sobre las posibles diferencias entre los grupos de edad considerados. Se sugiere también ampliar información sobre la selección y descripción de las muestras empleadas. Por un lado, se podrían especificar las mejoras en heterogeneidad y representatividad que se mencionan que se han incorporado en cuanto a nivel socioeconómico y cultural, ya que no se ofrecen detalles de estas características ni del tipo de muestreo realizado. Por otro lado, se podría añadir la distribución por sexo para cada baremo y el número de centros incluidos de cada tipo. Por último, y en relación a la distribución en función de la titularidad de los centros incluidos, se observa que el 24.9% del alumnado del conjunto de las muestras empleadas provienen de centros públicos. En caso que se compruebe que el tipo de centro se asocia a diferencias en las puntuaciones observadas en las diversas variables del test, sería conveniente comprobar si la distribución por tipo de centro en las muestras de baremación reproduce la del alumnado a nivel poblacional estatal.

En cuanto a la información presentada en los manuales, se proponen algunas mejoras. Podría optimizarse la organización del MTTC, diferenciando también más claramente la información de cada versión, para visibilizar mejor si se trata de información del conjunto de las pruebas o del nivel infantil BADyG/i, que en algunos aspectos parece ser la versión que más difiere de los restantes niveles. Existen discordancias en algunos valores numéricos entre el manual impreso y la versión digital que deberían subsanarse en una futura edición. También sería recomendable que si en los manuales no se adopta la terminología de los estándares más recientes citados por el propio COP en la versión del CET-R que complimentan lo/as revisores/as [AERA, APA & NCME. (1999, 2014)]. Standards for educational and psychological testing. Washington, DC: American Educational Research Association], se usen términos alternativos inequívocos. En este sentido, por ejemplo, resulta cuestionable la afirmación de los autores que ellos no deben probar directamente la validez por criterio de expertos, entre otras, y que un ejemplo de indicios de dicha validez es la evaluación que realiza el COP de tests que se usan en España (apartado 3.3.5. del manual técnico titulado "Otros indicios o sugerencias menos importantes de validación"; versión digital de 30/08/2019 consultada a 07/03/2020).

# ANÁLISIS DETALLADO DE LA PRUEBA

## 1. DESCRIPCIÓN GENERAL DEL TEST

**1.1. Nombre del test:**

BADyG/i: Bateria de Actividades mentales Diferenciales y Generales, Nivel infantil

**1.2. Nombre del test en su versión original:**

--

**1.3. Autor/es del test original:**

Carlos Yuste Hernanz y David Yuste Peña

**1.4. Autor/es de la adaptación española:**

--

**1.5. Editor del test en su versión original:**

CEPE (Ciencias de la Educación Preescolar y Especial)

**1.6. Editor de la adaptación española:**

--

**1.7. Fecha de publicación del test original:**

No se indica de forma específica (1980 para otros niveles)

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

--

**1.9. Fecha de la última revisión del test:**

No se indica de forma específica, se menciona la depuración de la versión experimental con datos de los cursos 1983-84 y 1984-85 y que la selección definitiva de los ítems fue hace 31 años. Los baremos actuales son de 2016 y la fecha del manual impreso revisado es 2019.

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

Inteligencia, Habilidades

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

El BADyG/i es un test de rendimiento máximo que evalúa inteligencia y habilidades mentales con 6 pruebas básicas como factores de primer orden (conceptos cuantitativo-numéricos, información, vocabulario gráfico, habilidad mental no-verbal, razonamiento con figuras y rompecabezas) y 2 pruebas complementarias (percepción auditiva y percepción y coordinación grafo-motriz); 2 aptitudes generales (inteligencia verbal e inteligencia no verbal) como factores de segundo orden; y un factor general de inteligencia general (IG) como factor de tercer orden. Además, ofrece puntuaciones de eficacia, omisiones y del cociente intelectual. Permite detectar los puntos fuertes y débiles del desarrollo cognitivo, y en caso de limitaciones en el aprendizaje poder orientar una consecuente intervención.

Pruebas básicas:

- Habilidad mental no-verbal (HMnV): prueba de razonamiento, de relacionar un conjunto de datos que configuran una escena con sentido. Consiste en seleccionar el dibujo que complete una escena.
- Conceptos Cuantitativo-Numéricos (CN): prueba de comprensión verbal de conceptos cuantitativo/numéricos básicos, que serán la base para futuros conocimientos aritméticos. Consiste en seleccionar el dibujo que cumpla el concepto presentado (contar, operar sumando o restando, comparar cantidades y/o magnitudes, discriminar numerales o símbolos aritméticos).
- Razonamiento con figuras (RF): prueba que mide la habilidad de clasificar. Consiste en seleccionar, de entre cinco dibujos, aquél que difiera más, al no presentar una característica esencial que, en cambio, permite englobar los restantes cuatro en un grupo.
- Información (Inf): prueba de asimilación cultural de conocimientos básicos del entorno social. Consiste en seleccionar el dibujo que cumpla una definición.
- Rompecabezas (Rpc): prueba perceptivo-espacial de comparación de figuras. Consiste en seleccionar la pieza que encaje en un dibujo al que se le ha recortado una parte de su superficie.
- Vocabulario gráfico (VG): prueba de asimilación significativa de conceptos y adquisición de vocabulario básico del entorno social. Consiste en seleccionar el dibujo que corresponda a un concepto (normalmente un sustantivo).

Pruebas complementarias:

- Percepción Auditiva (PA): prueba de discriminación auditiva de sonidos y su reproducción de manera automática. Consiste en repetir en voz alta una serie de palabras presentadas oralmente.
- Percepción y Coordinación Grafo-Motriz (PCGM): prueba de percepción visual y coordinación de los movimientos para reproducir configuraciones geométricas sencillas. Consiste en copiar un dibujo de tipo geométrico lo mejor posible.

Factores de segundo orden (a partir de las 6 pruebas básicas):

- Inteligencia General Verbal (IGV): incluye las 3 pruebas básicas con mayor contenido cultural, con un alto componente verbal y visoespacial y, en el caso del nivel infantil, también numérico, CN, Inf, VG.
- Inteligencia General no-Verbal (IGnV): incluye las 3 pruebas básicas con mayor contenido gráfico, HMnV, RF, Rpc.

Factor de tercer orden:

- Inteligencia general (IG): incluye las 6 pruebas básicas, es decir, los factores verbal y no verbal.



Otras puntuaciones adicionales:

- Eficacia (EF): % de aciertos , respecto a las respuestas emitidas (incluyendo el caso de doble marca) en las 6 pruebas básicas.
- Omisiones (OM): número de respuestas omitidas en las 6 pruebas básicas.
- Cociente intelectual (CI): a partir de la puntuación directa de IG teniendo en cuenta la edad cronológica de la persona evaluada.

#### 1.12. Áreas de aplicación:

Psicología Educativa

#### 1.13. Formato de los ítems:

Elección múltiple, respuesta construida

#### 1.14. Número de ítems:

Pruebas básicas:

- Habilidad mental no-verbal (HMnV): 18
- Conceptos cuantitativo-numéricos (CN): 18
- Razonamiento con figuras (RF): 18
- Información (Inf): 18
- Rompecabezas (Rpc): 18
- Vocabulario gráfico (VG): 18

Pruebas complementarias:

- Precepción auditiva (PA): 20
- Percepción y coordinación grafo-motriz (PCGM): 12

Total: 140

#### 1.15. Soporte:

Papel y lápiz, administración oral

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

Nivel B

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

Escolares de 2º ciclo de educación infantil y primer trimestre escolar de 1º de educación primaria (de 4 a 6 años y 11 meses).

Es aplicable a minorías étnicas en la medida en que se domine el idioma vehicular (en especial para las pruebas CN, Inf, VG) y haya integración cultural.

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

No

### 1.19. Procedimiento de corrección:

Manual, automatizada por ordenador (a través de Gestor BADyG on-line, previa introducción manual de las respuestas emitidas a cada elemento).

### 1.20. Puntuaciones:

Todos los ítems de las 6 pruebas básicas (HMnV, CN, RF, Inf, Rpc, VG) constan de varias opciones de respuesta, con una sola respuesta correcta. En la mayoría de casos el número de opciones es 5, excepto en la prueba CN en la que 12 de los 18 elementos tienen 4 opciones y 1 de ellos 3. En cada una de estas 6 pruebas básicas se otorga 1 punto por cada elemento correctamente contestado y 0 puntos cuando la respuesta es incorrecta, se ha omitido o hay doble marca. La puntuación directa de cada prueba básica es la suma de aciertos de los elementos (sin contabilizar los ejemplos), sin corrección por azar, por lo que puede oscilar entre 0 y 18.

En la prueba complementaria PA se otorga 1 punto por cada acierto y también se contabilizan los dos ejemplos, por lo que la puntuación directa oscila entre 0 y 20.

Y en la prueba complementaria PCGM se concede 0, 1 ó 2 puntos a cada elemento según se especifica en los criterios de corrección, por lo que la puntuación directa oscila entre 0 y 24.

La puntuación directa de los factores de segundo orden se obtienen por suma de las puntuaciones directas de las 3 pruebas básicas que los forman y oscilan entre 0 y 54:  $IGV = CN + Inf + VG$ ;  $IGnV = HMnV + RF + Rpc$

La puntuación directa del factor de tercer orden se obtiene por suma de las puntuaciones directas de las 6 pruebas básicas o, análogamente, por la suma de las puntuaciones directas de IGV e IGnV, y oscila entre 0 y 108.

La puntuación de eficacia (EF) se obtiene como el porcentaje en valor entero del número de aciertos (AC) en las 6 pruebas básicas (puntuación directa de IG) respecto a las respuestas emitidas (EM) y dobles marcas (DM) en las 6 pruebas básicas. Un valor alrededor del 25% puede indicar un patrón de respuestas al azar o gran dificultad para resolver las pruebas.

La puntuación de omisiones (OM) resulta del recuento de elementos sin respuesta de las 6 pruebas básicas.

El manual indica que en caso de duda sobre el dibujo marcado como respuesta (con una cruz) en las 6 pruebas básicas, se considerará el más cercano a la intersección de la cruz. En caso de doble marca se invalida la respuesta y si se dan 3 ó más dobles marcas en una prueba se considera nula. También se aconseja la administración individual si se observan respuestas estereotipadas o una prueba con cero aciertos. En la prueba complementaria PA, si tras tres repeticiones de los dos ejemplos no se han reproducido bien, se suspende la prueba.

### 1.21. Escalas utilizadas:

Centiles, Eneatipos, CI de desviación para la puntuación total.

### 1.22. Posibilidad de obtener informes automatizados:

Sí. La plataforma on-line (Gestor BADyG) para la corrección automática permite obtener 6 tipos de informes con un estilo claro, para detectar puntos fuertes y/o débiles tanto a nivel grupal como individual.

1) Colectivo: se detallan las puntuaciones directas y percentiles y valores de EF y OM de cada persona evaluada y los estadísticos descriptivos del grupo (n, M, DE)

2) Perfil individual: para cada persona evaluada seleccionada, se detallan las puntuaciones directas, así como una breve descripción de lo que evalúa cada escala, y los percentiles, que también se

muestran representados gráficamente mediante un diagrama de barras.

3) Descriptivo: para cada persona evaluada se incluye una explicación verbal individual según el nivel percentil obtenido en cada escala, resultando una valoración cualitativa del rendimiento general y en cada prueba, además de una apreciación pedagógica global de las pruebas básicas.

4) Comparativo: para cada prueba o factor se proporciona la media de las puntuaciones directas y del baremo de diferentes grupos de un determinado curso, resaltando en color los valores de media por debajo (rojo) o por encima (verde) de la media del baremo.

5) Distribución por niveles: para cada prueba o factor y para un determinado grupo, se muestra la distribución de las personas evaluadas entre 7 niveles en base a su percentil; también se muestra para qué pruebas la media del grupo es inferior (en rojo) o superior (en verde) a la media del baremo.

6) Valores de IG y CI y sus correspondientes percentiles: para cada valor individual observado se proporcionan los límites del intervalo de confianza de la puntuación verdadera.

Los informes 2 y 3, que son individuales, están dirigidos principalmente a las/os ma/padres de la persona evaluada; los informes 1, 4, 5 y 6, que incluyen información grupal o colectiva, están dirigidos a profesionales de la psicología y profesorado.

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

En aplicación individual: 1h (2-3 sesiones de 15-20 minutos, 2-3 pruebas por sesión con un descanso entre cada prueba).

En aplicación semicolectiva (grupos pequeños de entre 3 y 8 personas): mismo tiempo que en aplicación individual.

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

Manual, información técnica complementaria y actualizaciones (acceso al Manual Teórico-Técnico Completo MTTC a través del Gestor BADyG), cuadernos de elementos-respuestas

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

41,10€ sólo el manual técnico; 89,10€ (si se incluyen 60 usos para la corrección on-line) (18/01/2020)

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

2,71€ cada cuadernillo de elementos-respuestas (18/01/2020)

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

-- (incluido en el cuadernillo de elementos)

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

48,00€ licencia on-line de 60 usos para la corrección y/o elaboración de informes a través de Gestor BADyG on-line (0,80€ por participante)

149,00€ paquete de renovación de 300 usos

19,50€ paquete de renovación de 30 usos

(18/01/2020)

## 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	Buena	4
2.3. Fundamentación teórica	Excelente	5
2.4. Adaptación del test	-	-
2.5. Desarrollo de los ítems del test	Buena	4
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	Buena	4
2.10. Datos sobre el análisis de los ítems	Buenos	4

## 2.11. Validez:

### 2.11.1. Evidencias de validez de contenido:

Contenido	Valoración	Puntuación
2.11.1.1. Calidad de la representación del contenido o dominio	Buena	4
2.11.1.2. Consultas a expertos	-	-

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

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

Contenido	Valoración	Puntuación
2.11.2.1.1. Diseños empleados	Diferencias entre grupos (en función del sexo)	
2.11.2.1.2. Tamaño de las muestras	Un estudio con una muestra grande	3
2.11.2.1.3. Procedimiento de selección de las muestras	No se aporta información en la documentación	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	-	-
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	-	-
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	Buena	4

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

Contenido	Valoración	Puntuación
2.11.2.2.1. Criterios empleados	No se emplea ningún criterio	
2.11.2.2.2. Calidad de los criterios empleados	-	-
2.11.2.2.3. Relación temporal entre test y criterio	-	
2.11.2.2.4. Tamaño de las muestras	-	-
2.11.2.2.5. Procedimiento de selección de las muestras	-	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	-	-

2.11.3. Evidencias de validez basadas en la estructura interna:

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

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	<p>Sí.</p> <p>Variación del estilo tú a usted en aquellos contextos en que sea costumbre.</p> <p>Administración individual si se observan respuestas estereotipadas o una prueba con cero aciertos.</p>	

### 2.11.5. Comentarios generales sobre evidencias de validez:

Evidencias basadas en el contenido:

Se ofrece una muy amplia conceptualización de la inteligencia y de la definición del dominio y sus facetas. Sin embargo, no se menciona la consulta a personas expertas externas al equipo de autores para aportar evidencias empíricas en cuanto a la relevancia y la representatividad de los ítems y las referencias en el manual completo en formato digital son de tipo teórico. En cuanto al proceso de generación de los ítems y la depuración de la versión experimental inicial, en el manual impreso se alude a los restantes niveles del test, sin explicitar el proceso seguido en el caso del nivel infantil, aunque posteriormente se menciona que la selección definitiva de los elementos del nivel infantil está basada en el análisis cuantitativo de los ítems (índices de dificultad y discriminación).

En el manual completo en formato digital se aporta información de tipo teórico de los procesos requeridos para cada una de las pruebas básicas, lo que se correspondería con una evidencia basada en el proceso de respuesta. Para algunas de las pruebas básicas se mencionan numerosos instrumentos en los que se usan las mismas tareas como punto de partida para la confección de los elementos. No obstante, estos análisis de tareas comprenden las pruebas básicas de niveles superiores del BADyG, y no se incluyen específicamente para las pruebas del BADyG/i, que son diferentes debido al desarrollo cognitivo a los 4-6 años y siguen una lógica más icónica.

Evidencias basadas en la estructura interna:

Se aborda el estudio de la dimensionalidad del BADyG/i de forma exploratoria, mediante un Análisis en Componentes Principales (ACP), usando las puntuaciones de las 6 pruebas básicas como variables observadas. Las tablas de resultados mostradas parecen indicar que el programa empleado podría ser SPSS. La solución de dos componentes con rotación oblicua reproduce la estructura jerárquica hipotetizada de dos factores de segundo orden, Inteligencia General Verbal (IGV) e Inteligencia General no Verbal (IGnV) de forma moderadamente satisfactoria, ya que la matriz de configuración de las cargas factoriales rotadas no muestra una estructura simple perfecta. La prueba RF (Razonamiento con Figuras) satura en los dos componentes extraídos, pero en la documentación no se justifica que dicha puntuación sólo contribuya a la obtención de la puntuación del factor IGnV. El valor moderado-alto de la correlación entre los dos componentes permite inferir la presencia de un factor de tercer orden de Inteligencia General (IG).

Se valora positivamente que se hayan verificado las condiciones de aplicación del ACP mediante la medida KMO y la prueba de esfericidad de Bartlett. Sin embargo, y más allá de proporcionar los valores propios de los dos primeros componentes y el porcentaje de variancia explicada, se echa en falta una mayor justificación empírica del número de componentes extraídos (p.ej., gráfico de sedimentación, análisis paralelo) y del ajuste de la solución obtenida (p.ej., communalidades, % de residuales bajos).

Evidencias basadas en la relación con variables externas:

Se aportan evidencias basadas en la relación entre las puntuaciones del test y el sexo. Los resultados de las comparaciones van en el sentido esperado de no observar diferencias relevantes, lo que justifica el uso de baremos mixtos.

Respecto a evidencias basadas en la relación entre las puntuaciones del test y un criterio externo, en el manual completo en formato digital (consultado a 18/01/2020) se menciona que se proporcionan correlaciones con los resultados académicos al final del curso para aportar evidencias de carácter predictivo de los otros cinco niveles de las baterías BADyG, y para el nivel infantil ciertamente esta información no se ofrece, mientras que en el manual impreso (p. 57) consta que se hace para los seis niveles.

No se aportan evidencias de validez convergente y discriminante, que los autores dicen considerar menos necesarias en cuanto a importancia, interés y utilidad.

## 2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad. Coeficientes de fiabilidad para diferentes grupos de personas. Error típico de medida para diferentes grupos de personas	

### 2.12.2. Equivalencia formas paralelas:

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

### 2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Varios estudios con muestras grandes	5
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa Spearman-Brown, Guttman	
2.12.3.3. Promedio de coeficientes de consistencia	Adecuada	3

### 2.12.4. Estabilidad (test-retest):

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



### 2.12.5. Cuantificación de la precisión mediante TRI:

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

### 2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	No se aporta para las dos pruebas de respuesta construida (PA y PCGM)	
2.12.6.2. Promedio de los coeficientes	-	-

### 2.12.7. Comentarios generales sobre evidencias de fiabilidad:

El estudio de la fiabilidad de las puntuaciones del BADyG/i se centra en la consistencia interna mediante el coeficiente alfa de Cronbach de las puntuaciones de las 6 pruebas básicas y de los factores de orden superior, y el procedimiento de dos mitades (coeficiente de Spearman-Brown y Guttman) para los factores de segundo y tercer orden, en ambos casos por grupos de edad de 6 en 6 meses con un tamaño muestral grande (entre 651 y 9.409 escolares para las muestras de 2016).

Los valores de consistencia interna obtenidos pueden considerarse excelentes para el factor de tercer orden (promedio alfa de Cronbach ,90), buenos para los factores de segundo orden (promedio alfa ,84) y entre adecuados y con algunas carencias para las 6 pruebas básicas (promedio alfa de cada escala entre ,62 y ,75; promedio total ,68). El valor promedio para el conjunto de todas las medidas es de ,74. También se proporcionan los resultados del procedimiento de dos mitades obtenidos en una muestra de 158 escolares de 1985.

En el apartado de baremos se proporciona el error máximo, como producto del error estándar de medida multiplicado por el valor de z asociado al nivel de significación, para la estimación de los intervalos de confianza al 90% y 95% de las puntuaciones obtenidas (excepto para las 2 pruebas complementarias).

## 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	Muy grande	5
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.13.1.4. Procedimiento de selección de las muestras	No se aporta información en la documentación	
2.13.1.5. Actualización de baremos	Excelente	5

### 2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	-	-
2.13.2.2. Procedimiento empleado para fijar los puntos de corte	-	-
2.13.2.3. Procedimiento de obtención del acuerdo inter-jueces	-	-
2.13.2.4. Valor del coeficiente de acuerdo inter-jueces	-	-

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

La interpretación de las puntuaciones es normativa y se proporcionan 6 baremos mixtos diferenciados por edad de 6 en 6 meses, aunque en el manual se especifica que se obtienen en función del grupo o curso y no en función de la edad cronológica. Las muestras utilizadas para la baremación son en promedio muy grandes (entre 651 y 9.409 escolares; total de 21.940; promedio por baremo de más de 3.000 participantes), actuales (con datos actualizados en 2016) y proceden de centros públicos, concertados y privados de la mayoría de comunidades autónomas del territorio, aunque no se detalla el número de escuelas incluidas. Los baremos ofrecidos permiten transformar las puntuaciones directas en puntuaciones percentiles y eneatis; además para la puntuación directa de Inteligencia General se ofrece una estimación del Cociente Intelectual (CI con media 100 y desviación estándar 15) teniendo en cuenta la edad cronológica de la persona evaluada. También se proporcionan los estadísticos descriptivos de todas las medidas, que pueden permitir a las personas usuarias obtener otras puntuaciones estandarizadas derivadas.

En cuanto al uso de baremos mixtos, en el manual se justifica tanto a nivel teórico como empírico, ya que aunque para algunas variables las diferencias por sexo en una muestra de 16.018 escolares de 4 y 5 años son estadísticamente significativas, si se realiza el cálculo del tamaño del efecto se observa que dichas diferencias pueden considerarse irrelevantes ( $d$  de Cohen entre 0,00 y 0,10). No obstante, sería conveniente aclarar con qué datos se han hecho estas comparaciones, ya que el tamaño de la muestra empleada no coincide con el de las muestras para los baremos de dichas edades. En cambio, y aunque el uso de baremos diferenciados por edad pueda resultar lógico, no se proporciona información del criterio usado de 6 en 6 meses.

En uno de los informes (véase apartado 1.22, punto 5) se distribuyen los grupos en 7 niveles en base a los percentiles obtenidos, pero no se justifican los puntos de corte elegidos para establecer los diferentes estratos.