Consejo General de la Psicología
ESPAÑA

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

Batería de Actividades mentales Diferenciales y Generales, Nivel S, Renovado

BADyG/S-r

Prueba valorada en la ronda de evaluaciones del año 2019

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	BADyG/S-r: Batería de Actividades mentales Diferenciales y Generales, Nivel S, renovado
Autor	Carlos Yuste Herranz, David Yuste Peña, Rosario Martínez Arias, José Luis Galve
Autor de la adaptación española	-
Variable(s)	Inteligencia, habilidades
Áreas de aplicación	Psicología educativa y del trabajo y organizaciones
Soporte	Papel y lápiz, informatizado

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Buena-Excelente	4,5
Fundamentación teórica	Excelente	5
Adaptación	-	-
Análisis de ítems	Buena	4
Validez: contenido	EBuena - Excelente	4,5
Validez: relación con otras variables	Adecuada	3
Validez: estructura interna	Adecuada	3
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/S-r es un test de inteligencia construido a partir de la definición del constructo de Inteligencia General expuesto con detalle en el Manual Teórico-Técnico Completo (MTTC). Los autores definen la inteligencia como una actividad que requiere del pensamiento abstracto, que está incitada por la curiosidad innata y, que además, también permite predecir situaciones hipotéticas y anticipar las posibles consecuencias. Exponen una inteligencia jerarquizada y la evaluación de este constructo se lleva a cabo mediante 6 pruebas básicas: analogías verbales, series numéricas, matrices lógicas, completar oraciones, problemas numéricos y encajar figuras que se agrupan en un factor de inteligencia general y cuatro subfactores más específicos. Va orientado al ámbito educativo (1º y 2º de Bachillerato) y según se dice en el manual también puede aplicarse en procesos de selección de personal en niveles directivos aunque no se aportan datos sobre esta población.

Las evidencias de validez aportadas son adecuadas y los resultados son apropiados para la finalidad del test en el ámbito educativo.

La definición del dominio se explica de manera minuciosa y con detalle en el MTTC. Se echa en falta en la definición cuestiones como el ocio, la comunicación social, el aprendizaje escolar, el desarrollo profesional o las operaciones de introspección. Para la definición se han consultado tests prestigiosos que pretenden medir la inteligencia y se han tenido en cuenta las condiciones reales de administración. La calidad de los materiales del test (objetos, material impreso, software) es muy buena, de comprensión sencilla y empleabilidad fácil. Están bien especificadas las normas de aplicación e instrucciones, la población a la que va dirigida la prueba y la interpretación que se realiza de los resultados. Sería adecuado exponer con más detalle el proceso llevado a cabo para la creación de los ítems, además de la parte cuantitativa que los autores quieren seguir mejorando en las próximas revisiones, presentar el análisis cualitativo llevado a cabo por personas expertas, que se aconseja sean diferentes a las que desarrollaron los ítems.

Las evidencias de validez basadas en la estructura interna son adecuadas mediante dos análisis exploratorios en componentes principales, pero sería interesante realizar un estudio factorial confirmatorio con la finalidad de corroborar empíricamente los 4 subfactores que mide el test.

Por lo que respecta a las evidencias de validez en relación con otras variables, el manual muestra la relación con un criterio relevante, los resultados académicos. La comparación con otros tests similares debería ir acompañada de un análisis cuantitativo para poder considerarse como evidencia de validez convergente. Si no es así, recomendamos incluir esta discusión en el apartado de evidencias relativas al contenido o al proceso de respuesta. También podría llevarse a cabo un estudio cuantitativo para aportar evidencias de tipo discriminante para diferenciar cómo el test evalúa "Actividades mentales" y no otros constructos del ámbito de la inteligencia. Por otra parte, en el MTTC se presenta un estudio para valorar las diferencias de puntuación entre varones y mujeres cuyo resultado justifica la decisión de no presentar baremos diferenciados por sexo.

Respecto a los análisis realizados para el estudio de la fiabilidad, se observa un importante incremento en el tamaño de las muestras, respecto al primer análisis realizado en el año 2001. Los resultados obtenidos en el último análisis de consistencia interna son buenos. Los valores del coeficiente alfa de Cronbach son superiores a ,90 en inteligencia general, y moderados en el resto de escalas, en global, podemos valorar los resultados como buenos al presentar valores medios comprendidos entre ,80 y ,85. Aunque los resultados de consistencia interna son muy favorables y muestran una buena fiabilidad del test, el manual podría emplear otros índices con menos condiciones de aplicación como por ejemplo el coeficiente omega. Para el futuro, proponemos la posibilidad de estudiar la fiabilidad basada en la Teoría de Respuesta a los Ítems, puesto que el test ya cuenta con dos características favorables para tal propósito: disponer de los ítems ordenados por dificultad y la capacidad de recoger datos de muestras muy grandes. Por otro lado, los autores argumentan el no uso del test-retest como índice de estabilidad en la medida, apelando a que la inteligencia es un constructo dinámico, pero sin proponer ninguna alternativa para diversificar las evidencias de fiabilidad, como podría ser un estudio de medidas equivalentes. Realizar una prueba de dos mitades es un modo clásico y válido de aportar interesantes evidencias de fiabilidad al tratarse de dos pruebas completas aunque reducidas, pero sigue siendo evidencia basada en las interrelaciones entre los ítems del test.

Se ha realizado una actualización de los baremos incrementando el número de participantes

tanto de 1º como de 2º de Bachillerato, también ha aumentado la variabilidad de la procedencia (número de comunidades representadas) y se ha equilibrado por tipo de centro (público, privado y concertado). Por ello, concluimos que se ha aumentado la representatividad respecto a los baremos anteriores.

Es un factor a resaltar el hecho de que usen acomodaciones in situ para la correcta evaluación de personas con limitaciones, lo cual favorece la equidad y tiene en cuenta las posibles diferencias étnicas y/o culturales. Aun así, estas deberían estar especificadas previamente para que no recaiga toda la responsabilidad en el psicólogo/a que administra el test.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

BADyG/S-r: Batería de Actividades mentales Diferenciales y Generales, Nivel S, renovado

1.2. Nombre del test en su versión original:

-

1.3. Autor/es del test original:

Autor principal: Carlos Yuste Hernanz , y Confección estadísticas 2011: David Yuste Peña

Asesoría estadística versión renovada (2001): Rosario Martínez Arias. Colaboración especial: José Luis Galve Manzano

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:

1980, renovación 2001

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

-

1.9. Fecha de la última revisión del test:

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 test consta de 6 pruebas básicas: analogías verbales (Rv), series numéricas (Rn), matrices lógicas (Re), completar oraciones (Vv), problemas numéricos (Nn) y encajar figuras (Ge).

Estas pruebas básicas se agrupan en 4 estimaciones globales: razonamiento lógico (Rv, Rn, Re), factor verbal (Rv, Vv), factor numérico (Rn, Nn) y factor visioespacial (Re, Ge).

Estas estimaciones globales, a excepción de razonamiento lógico, dan lugar a la estimación de Inteligencia General (IG). A partir de la cual se puede obtener el coeficiente intelectual (CI). Por otra parte, este test también mide tanto la rapidez como la eficacia.

Por último, el test abarca la medición de algunos aspectos de la memoria auditiva (corto plazo) y visual-ortográfica (largo plazo) mediante 3 pruebas complemantarias: memoria de relato oral, memoria visual ortográfica, discriminación de diferencias.

Al finalizar el test, se pasa un cuestionario (autoinforme) que pretende estudiar la atribución del logro de la persona evaluada y detectar los posibles problemas circustanciales o emocionales que hayan podido darse durante la realización del test.

1.12. Áreas de aplicación:

Psicología Educativa y del Trabajo y las Organizaciones

1.13. Formato de los ítems:

Elección múltiple, respuesta abierta, respuesta dicotómica

1.14. Número de ítems:

El test consta de 6 pruebas básicas: analogías verbales (Rv), series numéricas (Rn), matrices lógicas (Re), completar oraciones (Vv), problemas numéricos (Nn) y encajar figuras (Ge).

Estas pruebas básicas se agrupan en 4 estimaciones globales: razonamiento lógico (Rv, Rn, Re), factor verbal (Rv, Vv), factor numérico (Rn, Nn) y factor visioespacial (Re, Ge).

Estas estimaciones globales, a excepción de razonamiento lógico, dan lugar a la estimación de Inteligencia General (IG). A partir de la cual se puede obtener el coeficiente intelectual (CI). Por otra parte, este test también mide tanto la rapidez como la eficacia.

Por último, el test abarca la medición de algunos aspectos de la memoria auditiva (corto plazo) y visual-ortográfica (largo plazo) mediante 3 pruebas complemantarias: memoria de relato oral, memoria visual ortográfica, discriminación de diferencias.

Al finalizar el test, se pasa un cuestionario (autoinforme) que pretende estudiar la atribución del logro de la persona evaluada y detectar los posibles problemas circustanciales o emocionales que hayan podido darse durante la realización del test.

1.15. Soporte:

Papel y lápiz, informatizado.

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

Nivel B

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

Este test va dirigido al alumnado de 1º y 2º de Bachillerato (16 - 20 años). También puede emplearse en

la selección de directivos/as (adultos con carrera universitaria), así como con al alumnado de ciclos formativos de grado superior. El test también puede dirigirse a grupos minoritarios pero con determinadas especificaciones. Concretamente, si va dirigido a una persona:

• que habla y entiende el español, pero que no es su lengua materna, se le podrán administrar 4 pruebas básicas: Rn, Nn, Re, Ge.

• recién llegada, que no domina el español, se le podrán aplicar 2 pruebas básicas: Re y Ge.

• perteneciente a ciertos grupos de integración y con alguna discapacidad funcional, será el/la psicólogo/a quien deba valorar si pueden participar en una administración colectiva..

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

Papel y lápiz, e informatizada (administración on-line que sólo necesita disponer de alguno de los siguientes navegadores: Internet Explorer 7.0 o superior; Firefox 3.0 o superior Firefox 3.0 o superior, Google Chrome o Safari).

1.19. Procedimiento de corrección:

Manual, mediante hoja autocorregible, por lectura óptica de la hoja de respuestas (COSPA), de forma automatizado por ordenador (gestor BADyG) y efectuada por la empresa suministradora.

1.20. Puntuaciones:

Las puntuaciones directas de las seis pruebas básicas es la suma de las respuestas correctas, a excepción de la prueba de Problemas Numéricos: Nn = sumatorio de las correctas - (sumatorio de las erróneas/2).

A partir de las puntuaciones de las pruebas básicas se pueden calcular las puntuaciones de los 4 factores globales. A partir de los factores globales, a excepción del razonamiento lógico, se obtiene la Inteligencia General (IG), y gracias a ésta, la puntuación del CI a través de una correlación entre las puntuaciones directas de IG y los baremos de edad de la persona evaluada.

La puntuación de rapidez es la suma de las respuestas emitidas y las omitidas.

La puntuación de eficacia se obtiene con la siguiente formula: (AC/ (EM+DM)) x 100, donde AC es el número de aciertos, EM la suma de las respuestas emitidas en las seis pruebas específicas y DM las dobles marcas.

También se calculan otras cualificaciones de la actividad mental, a parte de la eficacia (EF), como son la rapidez (RA), la atribución de logro y el cálculo de las omisiones (OM), se realiza simplemente sumando las respuestas omitidas en las seis pruebas básicas.

1.21. Escalas utilizadas:

Centiles, Eneatipos y Puntuaciones estandarizadas CI (Media 100 y desviación típica 15).

1.22. Posibilidad de obtener informes automatizados:

Sí, existe la posibilidad de obtener informes automatizados mediante la web www.gestorbadyg.com con la cual pueden obtenerse seis tipos de informes:

El "colectivo" muestra el listado completo del alumnado de un grupo con todas las puntuaciones directas y percentiles obtenidas en las distintas pruebas. También se muestra la media aritmética y la desviación típica tanto del curso como del grupo de baremación. En los listados aparecen abreviadas todas las escalas evaluadas y abajo se clarifica el nombre completo de cada escala. Este informe es útil para el/la psicólogo/a (principalmente) y para el/la profesor/a del grupo porque permite tener una visión global de todo el grupo.

El individual contiene la puntuación directa y percentiles en cada una de las pruebas con una representación gráfica. Este informe es útil para la persona evaluada y sus responsables legales, porque permite valorar los puntos fuertes y débiles de su inteligencia y de sus habilidades.

El descriptivo muestra una valoración textual de tres aspectos: las estimaciones globales, las pruebas aptitudinales básicas y las pruebas complementarias, añadiéndose por último, una apreciación pedagógica global de las pruebas básicas. Presenta la misma utilidad que el informe individual.

El comparativo de medias de las puntuaciones directas, tanto del baremo como de los distintos grupos del curso seleccionado. Cuando la media del grupo se encuentra por encima aparece subrayada por una línea verde y si queda por debajo aparece subrayada con una línea roja. Este informe es útil para el/la psicólogo/a y los profesores, ya que les permite identificar puntos fuertes y débiles de cada grupo y poder decidir futuras actuaciones a partir del resultado obtenido.

También está el informe de distribución donde se presenta cómo se reparte el alumnado de un curso entre los 7 niveles establecidos. Los niveles van desde muy alta hasta muy bajo. Además se indica con una línea roja en qué aptitudes la media del curso se encuentran por debajo de la media del baremo. Este informe es útil para el/la psicólogo/a y, especialmente, para el profesorado de asignaturas que incluyan actividades visoperceptivas, ya que proporcionan información sobre el ritmo que llevar durante el curso, en función del grupo y la materia que se imparte.

Finalmente, encontramos el informe de verdaderos valores de IG (inteligencia general) y CI en el que podemos ver los limites superiores e inferiores entre los que se encuentran estas puntuaciones para el alumnado de modo individual.

En general podemos observar que los informes son bastante claros y fáciles de entender, ya que nos dan la información de los baremos directamente y estos tienen una interpretación sencilla. Además algunos informes complementan la información aportando datos gráficos y colores que facilitan la interpretación.

1.23. Tiempo estimado para la aplicación del test:

Aplicación individual (lápiz y papel u online):

- Bachillerato: 1h y 29 minutos (65' pruebas básicas y 24' pruebas complementarias).

- Personas adultas (universitarias): 1h y 22 minutos (60' pruebas básicas y 22' pruebas complementarias).

El tiempo de aplicación colectiva es el mismo que el de aplicación individual.

1.24. Documentación aportada por el editor:

Manual impreso, información técnica complementaria (Manual Teórico Técnico Completo, MTTC, a través del Gestor BADyG), Cuadernos de elementos-preguntas, Cuadernos de respuestas y licencia de acceso al Gestor BADyG.

1.25. Precio de un juego completo de la prueba:

100,94€ (a 28/10/2019)

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

4,78€ (a 28/10/2019)

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

9,18€/paquete de 25 unidades (a 28/10/2019)

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

Corrección online: 0,80€ incluye la obtención de informes.

Administración y corrección online: 2,40€ (sin uso de cuadernillo ni hoja de respuesta). Incluye la obtención de informes.

El editor ofrece la posibilidad de la gestión de la corrección.

*** FE DE ERRORES EN PÁGINA 20**

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	Excelente	5
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 - Excelente	4,5
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	
2.11.2.1.2. Tamaño de las muestras	Un estudio con una muestra grande (N> 500)	3
2.11.2.1.3. Procedimiento de selección de las muestras	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	Rendimiento académico en lengua, matemáticas e inglés.	
2.11.2.2.2. Calidad de los criterios empleados	Buena	4
2.11.2.2.3. Relación temporal entre test y criterio	Predictivo	
2.11.2.2.4. Tamaño de las muestras	Un estudio con una muestra moderada (100 ≤ N <200)	2
2.11.2.2.5. Procedimiento de selección de las muestras	No se aporta información en la documentación	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Adecuada, pero con algunas carencias (0.20 ≤ r < 0.35)	2

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Adecuada	3
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	Sí	

2.11.5. Comentarios generales sobre evidencias de validez:

Las evidencias de validez aportadas son adecuadas y los resultados son apropiados para la finalidad del test, aunque podrían ampliarse en futuras ediciones del manual.

Las evidencias de validez de contenido en cuanto a la definición del dominio, se explica de manera minuciosa y con detalle en el Manual Teórico-Técnico Completo (MTTC). Además, los autores hacen referencia al uso de "Teorías psicométricas jerárquicas" para definir mejor el constructo, como: "La Teoría jerárquica de Cattell, la Teoría triárquica de Sternberg, les teorías de Cattell-Horn y la síntesis de Carroll". Estas aportaciones proporcionan robustas evidencias de validez de contenido al test. Se echa en falta en la definición cuestiones como el ocio, la comunicación social, el aprendizaje escolar, el desarrollo profesional o las operaciones de introspección. Todas ellas son cuestiones ya tomadas en cuenta por parte de las teorías actuales como la Teoría de la Inteligencia Múltiple, donde evalúan factores sociales o emocionales, por ejemplo. Los autores del test comentan que no contactaron con personas expertas para la definición del constructo, contar con su opinión aumentaría sustancialmente la calidad de las evidencias de contenido. También sería adecuado exponer con más detalle el proceso llevado a cabo para la creación de los ítems, y además de la parte cuantitativa que los autores quieren seguir mejorando en las próximas revisiones, presentar el análisis cualitativo llevado a cabo por personas expertas, que se aconseja sean diferentes a los que desarrollaron los ítems.

El lenguaje empleado al hacer referencia a la validez de constructo indica que se están empleando estándares ya abandonados por lo que se aconseja emplear los estándares más recientes [AERA, APA & NCME. (1999, 2014), Standards for educational and psychological testing. Washington, DC: American Educational Research Association], que denomina el tipo de evidencia que se presenta como validez relacionada con el contenido.

Respecto a las evidencias basadas en la estructura interna del test se destacan dos análisis factoriales exploratorios que aportan datos favorables al constructo de Inteligencia General. En primer lugar, el primer análisis de 2017, subdivide el test en pares e impares y realiza un análisis factorial con rotación promax, se obtienen 3 factores (numérico, verbal y espacial) que son las tres tipologías de contenido de las pruebas, pero no coinciden con los 4 subfactores que pretende estudiar el test. Aunque los resultados sean similares a los encontrados en estudios anteriores, sería adecuado incluir en la discusión una explicación convincente de por qué no se encuentra el factor de razonamiento. También debería incluirse una descripción de los criterios utilizados para determinar el número de factores así como el proceso seguido en este caso. Quizá sería interesante estudiar la posibilidad de realizar un análisis confirmatorio para verificar la hipótesis de partida expuesta al definir el dominio, donde se presentan los factores evaluados por el test: un factor global (inteligencia general) y cuatro subfactores. En segundo lugar se lleva cabo un análisis factorial con otro test, pasando a 225 alumnos/as, tanto el BADyG/S-r como el INVE/S (16 pruebas entre los dos), para, a continuación, realizar un análisis factorial con rotación varimax del que se obtienen 5 factores (verbal, memoria, razonamiento, lingüístico y visoespacial).

Los autores proporcionan también la comparación de las operaciones mentales que requieren los ítems del test con las operaciones mentales que se deben seguir en otros tests (Baterías de Actividades mentales Diferenciales y Generales, Baterías IGF-r, Baterías EDNIT). El análisis presentado es una valoración cualitativa y no un estudio cuantitativo de relaciones entre las puntuaciones de este test y los demás. Por este motivo debería evitarse calificarlo como un estudio de validez convergente. También podría llevarse a cabo un estudio para aportar evidencias de tipo discriminante especialmente para diferenciar cómo el test evalúa "Actividades mentales" y no otros constructos relativos a la inteligencia.

Respecto a las evidencias de validez relacionadas con un criterio relevante, el manual presenta un estudio útil y válido aunque modesto en lo que ha tamaño de la muestra se refiere. A principio de curso se les pasó el test a 182 alumnos/as de 1º de Bachillerato de la localidad de Guadalajara. El estudio ha confirmado la correlación positiva entre las puntuaciones del test al principio de curso con el rendimiento académico. Las correlaciones tenían de media valores entre ,20 y ,35, los cuales se considerarían valores adecuados, suficientes. La submuestra no parece ser representativa de la población diana, puesto que no se ha contado con alumnado de 2º de Bachillerato y por tanto los resultados no pueden generalizarse. Pero los resultados obtenidos en el sentido esperado alientan la realización de un estudio con una muestra más representativa a partir de la que poder generalizar los resultados.

Es importante destacar, aunque no se comente en el manual impreso (pero sí en el MTTC) el estudio que se realizó para valorar si había diferencias en las puntuaciones entre varones y mujeres. Es un estudio que se llevó a cabo en 2009 en todos los niveles del BADyG. La muestra fue de 57.908 estudiantes de los cuales a 2.977 se les administró el BADyG-S-r. De manera general, los autores concluyeron que realmente no existían diferencias claras, excepto a partir de los 18 años de edad.

En relación con las evidencias centradas en el proceso de respuesta, aunque el manual

impreso no especifique un apartado con este nombre, sí que aporta abundante información basada en el análisis de las tareas y descripción de procesos mentales necesarios para resolver las tareas. Todo de manera teórica, al comparar las operaciones mentales que requieren los ítems del test con las operaciones mentales que se deben seguir en otros tests (Baterías de Actividades mentales Diferenciales y Generales, Baterías IGF-r, Baterías EDNIT). A lo largo del manual se presentan diversas evidencias del proceso de respuesta desde el hecho de explicitar el tiempo necesario para la realización de las pruebas, pasando por contar con unas hojas de respuestas sencillas y fáciles de utilizar, hasta contar con un cuestionario final que nos informa sobre los posibles problemas que han podido surgir a la hora de contestar el test.

Y algo similar ocurre con las evidencias basadas en las consecuencias, pese a que los autores del manual no ofrecen un apartado específico, si cabe destacar que tienen en cuenta que la estimación de la habilidad intelectual que se hace con este test es una habilidad de ese momento y por ello nunca se deberían utilizar las puntuaciones con el fin de clasificar al alumnado. Podría resultar interesante que se valorarán de manera empírica las posibles repercusiones que puede generar a la persona evaluada la realización de estas pruebas y la obtención de las puntuaciones, teniendo en cuenta el constructo que se está evaluando y la importancia que se le da en nuestra sociedad. Finalmente, los autores tienen muy en cuenta los costes de la evaluación presentando los baremos en eneatipos para facilitar su comprensión, así como una variedad de informes para adaptarse a las múltiples necesidades de las personas implicadas.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad (para cada escala o subescala), 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 o KR-20	
2.12.3.3. Promedio de coeficientes de consistencia	Buena (,80 ≤ r < ,85)	4

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	-	
2.12.6.2. Promedio de los coeficientes	-	-

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

El estudio de la fiabilidad del test se ha llevado a cabo a través de la consistencia interna, mediante el coeficiente alfa de Cronbach y del procedimiento de las dos mitades (coeficiente de Spearman-Brown y de Guttman). Se obtuvieron datos de cada una de las pruebas, así como del factor general y factores globales, para cada uno de los cursos de la muestra (1º y 2º de Bachillerato).

En el estudio del 2017 se obtuvieron valores muy elevados en IG (,93 - ,94), alumnado de 1º y 2º respectivamente), valores similares a los encontrados en 2001 (,92 - ,95). Por lo que respecta a los coeficientes de fiabilidad de los factores globales en 2017 encontramos que en 1º de Bachillerato los valores oscilaban entre el ,83 y el ,88. Y de modo muy similar, entre ,83 y ,87 en 2º de Bachillerato. Por último, los coeficientes de consistencia interna de las pruebas básicas oscilan de modo muy similar en ambos cursos entre ,70 y ,84 (1º Bachillerato) y entre ,70 y ,86 (2º Bachillerato). En global, podemos valorar los resultados como buenos al presentar valores medios comprendidos entre ,80 y ,85. Con la finalidad de seguir mejorando el test, sería muy interesante en el futuro estudiar la fiabilidad de las diferentes combinaciones que recomienda el test para dirigirse a grupos minoritarios.

Siguiendo los criterios de Nunally y Bernstein, si el test se pretende usar para selección de personal y para hacer una interpretación de las puntuaciones individuales de gran impacto, el coeficiente alfa debería de ser igual o superior a ,90. En este caso, obtenemos esos valores en la puntuación de inteligencia general. El hecho de que no se alcance este nivel en las puntuaciones de los factores globales o de las pruebas básicas, no es problema, porque estos nunca se usan de modo individual sino que se emplean con otras pruebas básicas o factores globales. De todos modos, puesto que el test pretende también predecir el rendimiento académico del alumnado lo mejor es que las puntuaciones sean lo más precisas y con el menor grado de error posible.

Realizar una prueba de dos mitades es un modo clásico y válido de aportar interesantes evidencias de fiabilidad al tratarse de dos pruebas completas aunque reducidas. En este estudio quizá este método sea más adecuado que el coeficiente alfa de Cronbach, ya que es difícil que este test cumpla los requisitos para calcularlo. Por ello, sería útil aportar evidencias de fiabilidad complementarias, mediante un diseño de medidas equivalentes o bien aplicando la Teoría de Respuesta a los Ítems, puesto que el test ya cuenta con dos características favorables para tal propósito: disponer de los ítems ordenados por dificultad y la capacidad de recoger datos de muestras muy grandes.

También conviene matizar el comentario incluido en el manual sobre la estabilidad temporal basado en la estabilidad de los valores del alfa de Cronbach en el 2001 y 2017. Los coeficientes de consistencia interna cuantifican en qué medida hay concordancia entre las diferentes puntuaciones de las personas evaluadas en distintas partes del test. Que existan valores de alfa similares en dos momentos diferentes no asegura que las puntuaciones se mantengan estables a lo largo del tiempo.

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 poblaciones	4
2.13.1.2. Tamaño de las muestras	Muy grande (N ≥ 1000)	5
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	No se aporta información en la documentación	
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	-	-
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:

El manual proporciona baremos adecuados para dar una interpretación normativa a las puntuaciones de la población escolar.

Uno de los aspectos más relevantes a la hora de valorar los baremos es que se han llevado a cabo con una muestra muy grande y variada de la población, puesto que los baremos se hicieron a partir de una muestra de más de 4.000 alumnos/as de 1º (3.518) y 2º (581) de Bachillerato procedentes de centros públicos, privados y concertados. Los puntos débiles son la selección incidental de la muestra, y la distribución no equitativa por las comunidades autónomas y cursos académicos. En este sentido es muy visible que se seleccionaron muchos más estudiantes de 1º que de 2º de Bachillerato cuando en España, el porcentaje de estudiantes en 1º y de 2º de Bachillerato es muy similar.

Otra cuestión es que este test de inteligencia también va dirigido a estudiantes de ciclos formativos o adultos universitarios. En ningún momento se proporcionan baremos para la población universitaria o de ciclos formativos. Existe la posibilidad que los baremos resultaran similares a los de la población de bachillerato. No obstante, dado que no hay ninguna prueba de ello, sería interesante estudiarlo en un futuro.

En el estudio que se realizó para valorar las puntuaciones en función del sexo, los autores concluyen en el manual que se puede hacer uso de baremos mixtos porque no existen tales diferencias en el constructo evaluado. Esto contrasta con los resultados, presentados en el MTTC y obtenidos en función del sexo en la puntuación de Inteligencia General del BADyG/S-r, donde se encontraron diferencias significativas entre hombres y mujeres que los autores atribuyen a cuestiones culturales. Para futuras revisiones, se debería desarrollar esta argumentación, por ejemplo, apoyándose en el funcionamiento de otros test de inteligencia de reconocido prestigio.

No queremos terminar sin enfatizar de nuevo que la baremación que se presenta conlleva un trabajo muy costoso. Aunque sería deseable, no es frecuente encontrar baremos basados en muestras tan grandes.

*** FE DE ERRORES**

Con fecha 19 de mayo de 2020 se han modificado el epígrafe 1.28 incluido en la primera versión del documento, página 11, con objeto de subsanar un error en la política de precios ajeno a la empresa editora del test.