

Evaluación de

Batería de Actividades mentales Diferenciales y Generales, Nivel E2 Renovado

 **BADyG/E2-r** 

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Bateria de Actividades Mentales Diferenciales y Generales Nivel E2 renovado (BADyG-E2r)
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, informatizado y administración oral

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Buena - Excelente	4,5
Fundamentación teórica	Excelente	5
Adaptación	-	-
Análisis de ítems	Buena	4
Validez: contenido	Excelente	5
Validez: relación con otras variables	Adecuada con algunas carencias - Adecuada	2,7
Validez: estructura interna	Adecuada	3
Validez: análisis del DIF	-	-
Fiabilidad: equivalencia	-	-
Fiabilidad: consistencia interna	Excelente	5
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-E2r es una batería de tests cuyo objetivo es evaluar las actividades mentales diferenciales y generales de la inteligencia en estudiantes de tercero y cuarto de Educación Primaria Obligatoria (EPO). Se compone por seis pruebas básicas y tres complementarias, cuyas puntuaciones se agrupan en cuatro factores de segundo orden, un factor de tercer orden (inteligencia general) y tres puntuaciones extras (rapidez, eficacia y omisiones). El test es muy utilizado y apreciado para orientación en el ambiente escolar, cuenta con diferentes revisiones y baremos actualizados y es de fácil manejo y comprensión.

La calidad de los materiales del test es muy buena, tiene unas instrucciones muy claras, una buena explicación sobre el desarrollo de los ítems, y una clara y fácil manera de registrar las respuestas. Puede realizarse una administración a papel y lápiz donde la persona que administra debe manejar con soltura y precisión el manual técnico y el estudiante dispone de un cuadernillo de preguntas y otro de respuestas. También existe la posibilidad de realizar una aplicación online que facilita la corrección y reduce la posibilidad de cometer errores en las respuestas. En el primer caso se utilizan plantillas individuales para corregir en el segundo caso a partir de la plataforma online se puede hacer una administración y corrección de forma rápida y útil.

En referencia a las evidencias de validez, tanto el manual técnico como el manual teórico-técnico-completo (MTTC) ofrecen una gran fundamentación teórica aportando pues evidencias de validez basadas en el contenido. También cabe destacar el minucioso análisis de tareas, viendo el proceso de respuesta que debería seguir una persona al contestar. Respecto a las evidencias de validez basadas en la estructura interna del cuestionario, se aporta evidencia a partir de dos estudios que utilizan análisis de componentes principales: el primer estudio quiere buscar un factor de inteligencia global; el segundo estudio quiere buscar tres factores de segundo orden.

Uno de los posibles usos del test es realizar el pronóstico de los rendimientos académicos de los estudiantes. Para validar este uso, se establecen dos criterios externos: la puntuación que el profesorado de los estudiantes dan de matemáticas y lengua. En general, las correlaciones entre las variables del test y los rendimientos son positivas y estadísticamente significativas tal como era de esperar.

Las evidencias aportadas en relación con la fiabilidad de las puntuaciones del test son a partir del análisis del coeficiente alfa de Cronbach así como los que coeficientes de Spearman-Brown y Guttman de dos mitades. Para alfa de Cronbach, se aportan datos relativos a 2001 y 2014. Los resultados muestran valores superiores a ,90 en todos los casos para inteligencia general, entre ,85 y ,93 para factores de segundo orden, y una fiabilidad alrededor de ,80 para todas las pruebas básicas y complementarias.

De las evidencias aportadas, cabe destacar algunas limitaciones o puntos que podrían mejorar. En primer lugar, sería idóneo introducir un cierto orden en las fechas e incluso hacer algunas correcciones de redacción. Por ejemplo, prólogo (pg. 9) último párrafo. En la misma línea, se podría plantear la simplificación de la información contenida en el MTTC ya que en ocasiones resulta de difícil comprensión debido a la gran cantidad de información que se presenta.

Por otro lado en dos ocasiones (correlación del test con el criterio y alfa de Cronbach de 2001) se indica que no se muestran tres factores de segundo orden, justificando su ausencia debido a limitaciones en la recogida de datos previa. Aunque en 2014 sí que se calculó la fiabilidad de las puntuaciones de estos factores, no se aporta evidencia de relación con otras variables (criterio) para esas dimensiones.

Finalmente, y teniendo en cuenta las técnicas escogidas para obtener evidencias de validez y fiabilidad, también se podrían mejorar algunos aspectos. En primer lugar, se echa en falta evidencia de validez en relación con otras variables ya sea convergente y/o discriminante. En segundo lugar, la información sobre fiabilidad se diversificaría si se aportaran los valores obtenidos con otros diseños no centrados en el estudio de la consistencia interna. Además, la elección de la técnica para examinar la estructura interna del test podría reconsiderarse, sobre todo teniendo en cuenta que este test fue desarrollada hace 40 años y refleja una base teórica consolidada. Se podría pues aportar evidencia a partir de análisis factorial confirmatorio o también exploratorio aportando índices de ajuste (tipo ESEM), o incluso plantearse hacer el análisis desde la perspectiva de la teoría de respuesta al ítem que facilitaría tener en cuenta que los ítems utilizados tienen diferente dificultad.

En definitiva, el BADyD/E2-r es una buena batería de actividades para valorar las habilidades mentales diferenciales y básicas para estudiantes de tercero y cuarto de la EPO. Su administración es fácil y la interpretación de resultados completa y útil para el sistema escolar. Se recogen con toda precisión las limitaciones de la prueba para determinados grupos minoritarios. Deja a juicio de la profesional que lo aplica la valoración de las limitaciones funcionales que podrían tener algunos estudiantes para utilizar algunas de las pruebas de la batería.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Batería de Actividades Mentales Diferenciales y Generales Nivel E2 renovado (BADyG/E2r)

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:

1980

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

-

1.9. Fecha de la última revisión del test:

2019. Renovación de los ítems en 2001. Entre 2001 y 2014 se amplía la fundamentación teórica y se obtienen nuevas muestras para obtener evidencias de fiabilidad y validez. Se hace una revisión en 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-E2r pretende discriminar habilidades mentales competenciales y obtener estimaciones de actividades intelectuales más generales. El test está basado en un enfoque multidimensional a partir de seis escalas de primer nivel (básicas) y tres complementarias, que se agrupan en cuatro factores de segundo nivel, los cuales permiten obtener la puntuación de inteligencia general y de eficacia y rapidez. En referencia a las seis pruebas básicas se distingue Relaciones analógicas (Rv), Problemas numérico-verbales (Rn), Matrices lógicas (Re), Completar oraciones (Vv), Cálculo numérico (Nn) y Figuras giradas (Ge):

1. Relaciones Analógicas (Rv): aptitud para establecer relaciones analógicas verbales. Prueba de razonamiento y comprensión verbal. Implica reconocimiento de significados, y relaciones de segundo orden entre ellos, codificación verbal, pensamiento abstracto, memoria a largo plazo y atención.
2. Problemas Numérico-Verbales (Rn): evalúa la flexibilidad y rapidez para resolver problemas numérico-verbales, que exigen realizar cálculos sencillos y entender el planteamiento de conceptos aritméticos. Exige razonamiento inductivo, cálculo, memoria de trabajo y atención.
3. Matrices Lógicas (Re): Evalúa la habilidad para encontrar leyes inductivas lógicas en la organización de figuras geométricas. Exige abstracción y percepción, memoria de trabajo y atención.
4. Completar Oraciones (Vv): Evalúa la habilidad para codificar y elegir la palabra que completa el hueco de la oración. Requiere reconocimiento de significados integrados de situaciones que exigen tener unos conocimientos previos. Implica razonamiento analógico, comprensión, cierre de significado, codificación, memoria a largo plazo y atención.
5. Cálculo numérico (Nn): Evalúa la agilidad y rapidez mental para realizar cálculos numéricos sencillos (sumar y restar). Exige comprensión, memoria a largo plazo y atención.
6. Figuras Giradas (Ge): Evalúa la flexibilidad para rotar figuras geométricas, manteniendo sus relaciones de tamaño, distancia, forma y posición relativa. Exige comparación (diferencias- semejanzas), memoria y atención.

Para las pruebas complementarias se distingue entre Memoria de relato oral (Ma), Memoria visual ortográfica (Mv) y Discriminación de diferencias (De):

7. Memoria de Relato oral (Ma): Evalúa la habilidad para recordar a corto plazo datos de un relato verbal. Prueba verbal y numérica por los contenidos a recordar. Implica memoria inmediata auditiva y atención.
8. Memoria Visual Ortográfica (Mv): Evalúa la retención en la memoria a largo plazo de la escritura correcta de las palabras, que fonéticamente se pronuncian igual, pero se escriben de manera distinta. Exige selección perceptiva para elegir la palabra que está ortográficamente está bien escrita.
- 9 Discriminación de Diferencias (De): Evalúa la habilidad para discriminar, con rapidez y precisión, pequeñas diferencias visuales entre figuras y dibujos sencillos.

La variable de Inteligencia General (IG) describe actividades de operaciones de codificación de información, de elaboración inductiva, de comprensión significativa, de memorización, retención y evocación de la información. Se obtiene por tanto, a partir de las de segundo orden: VV (factor verbal) NN (factor numérico) y EE (factor visoespacial), las cuales a su vez, se obtienen a partir de seis pruebas básicas. A estas escalas de segundo orden se le añade también el factor RR (razonamiento lógico). En el factor numérico (NN) se presentan diferentes pruebas de contenido general numérico. El factor verbal (VV) contiene pruebas de contenido verbal, entre las que se mezclan contenidos de comprensión verbal y razonamiento analógico. Por último, en el factor visoespacial (EE) encontramos pruebas que requieren habilidades referentes a representaciones bidimensionales, entre otras. En el

factor de razonamiento lógico (RR) se integran operaciones inductivas y conceptos más abstractos con contenido visoespacial, numérico y verbal.

1.12. **Áreas de aplicación:**

Psicología educativa

1.13. **Formato de los ítems:**

Respuesta construida y elección múltiple.

1.14. **Número de ítems:**

El BADyG-E2r, consta de un total de 216 ítems. Cada una de las nueve pruebas consta de 24 ítems. Existen seis pruebas específicas (144 ítems) y tres complementarias (72 ítems).

1.15. **Soporte:**

Papel y lápiz e informatizado. No obstante, para una de las pruebas es necesaria la 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:**

El test está dirigido a estudiantes de tercero y cuarto de EPO con edades comprendidas entre los ocho y los 10 años. En el manual se describe la posibilidad de aplicarlo a poblaciones minoritarias, aunque se desconocen las propiedades psicométricas de las puntuaciones del test en estos casos: estudiantes que no tienen el español como lengua materna, poblaciones de inmigrantes y grupos con limitaciones funcionales.

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

Existe la versión impresa en papel, y más recientemente la versión informatizada.

1.19. **Procedimiento de corrección:**

Manual, hoja autocorregible, lectura óptica de la hoja de respuestas, automatizado por ordenador, así como otras empresas que ofrecen servicio de corrección (CEPE, COSPA y DINTEST)

1.20. **Puntuaciones:**

Para todas las pruebas, excepto para Cálculo numérico (Nn), la puntuación directa es la suma del número de aciertos. Para Nn, a la suma del número de aciertos se le restará también el número de errores. Por lo tanto, para todas las pruebas la puntuación directa máxima es de 24 puntos.

Para obtener las puntuaciones globales se suman las puntuaciones directas (PD) de las pruebas básicas que conforman los factores de segundo y tercer orden:

-Razonamiento Lógico (RR)=Rv+Rn+Re: PD Máxima 72

-Verbal (VV)= Rv+Vv= PD: Máxima: 48 (error manual dice 64, pg. 36)

-Numérico (NN)= Rn+Nn: PD. Máxima 48 (error manual dice 64, pg. 36)

-Visoespacial (EE)= Re+Ge: PD Máxima 48 (error manual dice 64, pg. 36)

-Inteligencia General (IG)= $R_v+R_n+R_e+V_v+N_n+G_e$: PD Máxima 144. A partir de esta puntuación directa también es posible obtener el cociente intelectual (CI) considerando la edad cronológica (EC) con el Gestor BADyG.

Además, se pueden obtener también puntuaciones de omisiones, rapidez y eficacia:

-Omisión (OM): se obtiene a partir del número de respuestas omitidas en las seis pruebas

-Rapidez (RA): Respuestas emitidas (EM) +Respuestas Omitidas (OM). En las 6 pruebas básicas. Total 144

-Eficacia (EF)= $(AC / (EM+DM)) \times 100$. Donde, AC= número de aciertos o puntuación directa de la IG. DM= número de dobles marcas (las preguntas que se invalidan). La PD máxima es 100.

1.21. Escalas utilizadas:

Centiles, eneatispos, CI de desviación (Media 100 y desviación típica 15, Weschler) y puntuaciones estandarizadas

1.22. Posibilidad de obtener informes automatizados:

Hay la posibilidad de obtener hasta seis tipos de informes automatizados. En primer lugar hay un informe individual que muestra el perfil de cada estudiante. En segundo lugar, también existe un informe colectivo que describe los estadísticos del grupo-sección incluyendo nombres de las personas y aplicando los baremos correspondientes. También existe la posibilidad de obtener un informe descriptivo en el cual se ven los resultados redactados destacando los puntos fuertes y las debilidades de las personas. Otro tipo de informe es uno que resulta de la comparativa de medias del grupo con las de los baremos. En la misma línea también existe otro informe que permite situar a cada estudiante en siete niveles de muy bajo hasta muy alto, ayudando a diseñar e implementar programas y acciones para paliar lagunas y fortalecer puntos fuertes. Finalmente, también existe otro informe en el que se obtienen los verdaderos valores de IG y CI.

1.23. Tiempo estimado para la aplicación del test:

Tanto para aplicación individual o colectiva se estima al mismo tiempo en función del curso en el que están los estudiantes. Para tercero de primaria se estiman 84 minutos: 62 minutos para las pruebas básicas y 22 minutos para las pruebas complementarias. Para cuarto de primaria se estima en 76 minutos: 56 minutos para las pruebas básicas y 20 minutos para las pruebas complementarias.

1.24. Documentación aportada por el editor:

El editor aporta el manual, información técnica complementaria y actualizaciones, así como licencia de acceso al gestor BADyG.

1.25. Precio de un juego completo de la prueba:

Manual Técnico: 41,10€.

1.Papel y lápiz : Cuadernillo (3,80€)+Hoja de respuesta (1,46€)= 5,26 €

2. Online: Gestor licencia con 60 usos = 48,00 € (renovación 30 usos: 19,50 €, renovación 300 usos: 149 €)

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

3,80€ c/u

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

1,46€ c/u

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

19,50€ por 30 usos o 149,00€ por 150 usos

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	Buena	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	Excelente	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	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 escolar (Matemáticas y lengua)	
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 o varios estudios con muestras pequeñas (N < 100)	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	3

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:

En la documentación aportada por los editores, se dan múltiples evidencias de validez de las puntuaciones del BADyG-E2r. Un punto fuerte de este son las evidencias basadas en el contenido. Este test, fue desarrollado a partir de una gran fundamentación teórica siguiendo las teorías psicométricas jerárquicas de la inteligencia. En el manual se aporta una amplia revisión en torno a las teorías en las que se basan las diferentes pruebas, y además en el manual teórico-técnico-completo (MTTC) se amplían estas explicaciones. Los autores también comentan que algunas de estas pruebas fueron desarrolladas a partir de cuestionarios ya previamente existentes como puede ser el test de matrices de Raven.

Mediante un análisis de tareas en el manual se dan evidencias tanto basadas en el contenido como en el proceso de respuesta. Se analizan con toda precisión las seis pruebas básicas (Rv, Vv, Rn, Nn, Re y De), según la situación del problema, los objetivos, subobjetivos y las estrategias ejecutivas que se requieren para su solución. Para esta labor se han analizado pruebas clásicas de inteligencia donde ya se incluían contenidos y operaciones mentales. Cabe destacar que en el manual a este apartado se le denomina validez concurrente, pudiendo inducir erróneamente al lector a pensar que el apartado trata sobre la relación entre los resultados de las pruebas.

Las evidencias de validez basadas en la estructura interna del test se encuentran en el apartado del manual dedicado al estudio de la consistencia interna y se derivan del análisis en componentes principales (ACP). En un primer análisis, en 2001 con una muestra de 480 estudiantes de cuarto de EPO buscan una estructura unifactorial para demostrar la existencia de un factor general de inteligencia a partir de la puntuación directa de las 6 pruebas básicas. En un segundo estudio realizado alrededor de 2015 con una muestra conjunta de 4.067 estudiantes de tercero y cuarto de EPO se examinan los factores de segundo orden especificados en la teoría (Factor numérico, factor espacial, y factor verbal). Para ello recurren a la parcelación de cada una de las pruebas básicas en ítems pares e impares y se realiza el análisis por separado en tercero y en cuarto de EPO. El debate sobre el uso de la técnica de parcelación podría evitarse utilizando otras técnicas de análisis de la estructura interna como por ejemplo la teoría de respuesta a los ítems.

Finalmente, se ofrecen evidencias basadas en la relación entre las puntuaciones del test y otras variables, en concreto se realizan dos estudios. El primero muestra la relación del test con un criterio específico: la predicción del rendimiento escolar (evaluado por el profesorado a cargo los estudiantes que responden el test y del cual se obtienen las puntuaciones predictoras del rendimiento escolar). A pesar de ser una muestra moderada (137 participantes) y poco representativa (todos los participantes de un único curso de la misma región), existen correlaciones entre las puntuaciones del test los dos criterios escogidos siendo una excepción las puntuaciones de eficacia y rapidez. El segundo estudio, desarrollado en el MTTC compara dos grupos distintos (sexo masculino vs sexo femenino). Las escasas diferencias encontradas en este estudio apoyan la decisión de no presentar baremos por separado entre sexo masculino y femenino para este test.

En el tercer capítulo del MTTC "análisis de elementos, fiabilidad, validez" se menciona un apartado de "otros indicios o sugerencias menos importantes de validación" que incluyen validez aparente, validez por criterio de expertos, y validez divergente. Sin embargo, teniendo en cuenta las directrices actuales en psicometría, convendría aportar también estos tipos de evidencias de validez para complementar las que ya se aportan.

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 y 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 de Cronbach	
2.12.3.3. Promedio de coeficientes de consistencia	Excelente	5

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:

Sobre las evidencias de fiabilidad, se aporta evidencia a partir del estudio de la consistencia interna en dos muestras. La rúbrica para la evaluación de tests califica como excelentes los valores de consistencia interna superiores a ,85. Sin embargo, los autores determinan un grado de fiabilidad mínimo para los factores teniendo en cuenta que se trata de un test de inteligencia. Para el factor de inteligencia general fijan un mínimo de ,90, para los de segundo orden requieren valores entre ,80 y ,90 y para los de primer orden una fiabilidad aproximada de ,80.

En primer lugar, a partir de una muestra del año 2001 (n = 429 en tercero de EPO y n = 718 en cuarto de EPO) en los cuales se aporta el valor del coeficiente alfa de Cronbach para todos los factores excepto para los factores de segundo orden EE, VV, y NN. Los datos aportados cumplen los requisitos que los autores se habían marcado.

En segundo lugar a partir de una muestra del año 2014 (n = 1152 en tercero de EPO y n = 962 en cuarto de EPO) se estudia la fiabilidad de consistencia interna mediante alfa de Cronbach cumpliendo también con los requisitos que los autores se habían marcado. Además, también se examina la fiabilidad de dos mitades (coeficiente de Guttman y coeficiente de Spearman Brown), obteniendo también resultados satisfactorios. Además, también se aporta el error máximo tanto para tercero como para cuarto de EPO.

Por último, cabe destacar que no se ofrecen evidencias referentes a la estabilidad temporal del test (test-retest), dado que los autores consideran que este tipo de evidencia presenta demasiados problemas teóricos, entre los que mencionan el hecho que estos estudios implican aceptar que el sujeto no cambia en el tiempo, la incógnita que representa saber cuánto tiempo dejar entre administraciones o el hecho que son estudios poco económicos. No obstante, indican que en ediciones anteriores del test habían obtenido este coeficiente, mostrando resultados similares a los de las dos mitades. Considerando que ambos coeficientes miden aspectos diferentes de la fiabilidad sería adecuado considerar la estabilidad temporal, o al menos aportar los resultados de este estudio previo aunque fuera con otros reactivos.

En resumen, la fiabilidad medida con el coeficiente alfa de Cronbach y con el método de dos mitades (Spearman-Brown, Guttman) es muy elevada. Si observamos los resultados obtenidos con las muestras de ambos años y para ambos cursos de alfa de Cronbach, y del 2014 de dos mitades, encontramos los coeficientes de fiabilidad siguientes

en el factor principal y los factores de segundo orden: IG ($\alpha = ,94-.96$, dos mitades = $,93$), RR ($\alpha = ,90-.94$, dos mitades = $,96$), VV ($\alpha = ,85-.87$, dos mitades = $,88-.89$), NN ($\alpha = ,92-.93$, dos mitades = $,88-.89$) y EE ($\alpha = ,85-.87$, dos mitades = $,85-.88$) y una fiabilidad alrededor de $,80$ para todas las pruebas básicas y complementarias. Por lo tanto, mediante las evidencias aportadas podemos decir que la fiabilidad de consistencia interna de las puntuaciones de ese test es excelente.

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	
2.12.1.4. Procedimiento de selección de las muestras	No se aporta información en la documentación	
2.12.1.5. 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:

Los baremos han sido actualizados con datos recogidos entre 2011 y 2015, ampliando el número de estudiantes. Según el manual estas muestras son más representativas que las anteriores ya que incluyen centros públicos concertados y privados. No obstante falta información sobre el proceso de selección de las muestras para evaluar con precisión si realmente es una muestra representativa o no.

Los baremos presentados están divididos en función del grupo o curso académico ($n = 1.152$ en tercero de EPO y $n = 962$ en cuarto de EPO) y no en función de la edad cronológica (excepto el CI). Se reconoce que algunas de las pruebas no serán adecuadas para los estudiantes que no dominen la lengua española, especialmente las pruebas de dominio verbal (Rv, Vv y Ma) aunque no se indica cómo proceder en estos casos.

Los baremos proporcionan la conversión de puntuaciones directas a eneatis y percentiles para cada una de las diferentes escalas. Como se ha avanzado, a partir de la corrección online, también se ofrece la posibilidad de realizar una estimación del CI a partir de la puntuación directa en inteligencia general y la edad, eliminando así el sesgo por edad. Como se ha mencionado en el apartado de fiabilidad, también se ofrece el error máximo que permite el cálculo de la puntuación verdadera de los estudiantes.

Cabe destacar la importancia de que estos baremos sean actualizados. Como los autores del manual indican, respecto a los baremos anteriores (2001) la puntuación media de todas las pruebas básicas ha aumentado y en consecuencia también las puntuaciones referidas a cada percentil en los baremos. Ello puede ser un reflejo del efecto Flynn en la medida de la inteligencia.