

Evaluación de

Batería de Actividades mentales Diferenciales y Generales, Nivel M renovado

 **BADyG/M-r** 

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	ADyG/M-r: Batería de Actividades mentales Diferenciales y Generales, Nivel M, renovado
Autor	Carlos Yuste, David Yuste, Rosario Martínez Arias y José Luis Galve Manzano
Autor de la adaptación española	-
Variable(s)	Inteligencia, habilidades
Áreas de aplicación	Psicología educativa
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	Bueno	4
Validez: contenido	Excelente	5
Validez: relación con otras variables	Adecuada - Buena	3,7
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	Bueno - Excelente	4,7

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

Comentarios generales

La Batería BADyG M-R mide la inteligencia a través de la realización de actividades mentales en las que entran en juego principalmente procesos de razonamiento y memorísticos aplicados a tres tipos de códigos (verbal, numérico y visoespacial) que evolucionan de manera continua a lo largo de la vida escolar de las personas. Su objetivo es evaluar su nivel intelectual a estudiante de los niveles de 1º a 4º de ESO que dominen el castellano y sin problemas funcionales graves y detectar aquellos casos con valores excesivamente bajos proponiendo estrategias de intervención para su mejora. También presenta como objetivo la selección de personal en niveles administrativos, aunque no se encuentra justificación psicométrica en el manual para esta aplicación.

Respecto a las versiones anteriores presenta la actualización de los baremos cuyas muestra normativas han mejorado en precisión y representatividad al contar con un mayor número de personas y de características más variadas, y una mejora del gestor 'on-line' que permite responder el test informáticamente y/o la corrección automatizada de las pruebas con el cálculo y transformación de las puntuaciones así como la obtención de informes exhaustivos. El mismo gestor sirve como repositorio de toda la información técnica complementaria al manual facilitado.

El Test permite la obtención de una puntuación general de Inteligencia a partir de tres factores: Verbal, Numérico y Viso-espacial, así como una puntuación en razonamiento y el cálculo de un CI para cada uno de los evaluados. A partir de estos resultados se elaboran los perfiles descriptivos contando con una gran variedad de resultados y unas instrucciones de interpretación claras y precisas.

Dichas interpretaciones vienen apoyadas por numerosas evidencias teóricas respecto a la definición del dominio y a los procesos de respuesta que son necesarios aplicar para la resolución de los ítems, todas ellas de un nivel excelente. También aporta evidencias excelentes sobre la relación de las puntuaciones con el rendimiento escolar (criterial predictiva) y con el sexo, comprobándose las hipótesis de partida de no diferencias entre los grupos. Menos agraciados son el intento de sustituir el test por una forma calificada como 'casi paralela' para construir una evidencia de validez convergente que, por dicha razón, pierde toda su utilidad, y la falta de evidencias de validez discriminante.

Respecto a la estructura interna, presenta análisis exploratorios en componentes principales que apoyan las especificaciones expuestas en la definición del dominio, obteniéndose por un lado un factor general de inteligencia, y por otro, los tres factores correlacionados correspondientes a los códigos verbal, numérico y viso-espacial (el factor razonamiento se deduce transversalmente de estos tres factores) siempre partiendo de las puntuación de las subescalas, no de los ítems. En este aspecto se echan de menos análisis confirmatorios que refuercen las evidencias ya presentadas.

Los resultados de fiabilidad son también excelentes, obteniéndose coeficientes con los diseños de dos mitades corregido y de alfa de Cronbach, pese a que los ítems no son equivalentes ya que presentan diferentes grados de dificultad. En este caso quizás sería más interesante la aplicación de formas paralelas bien diseñadas, o de técnicas basadas en la Teoría de Respuesta a los Ítems. Estos índices son considerados muy buenos por la mayoría de expertos y aptos para calificaciones individuales en las escalas más generales. Se acompañan además por los errores de medida (citados como $D_{máx}$) que dan una idea más práctica de la buena precisión de las puntuaciones obtenidas.

En cuanto a los aspectos formales del test, los materiales de los elementos facilitados son de alta calidad, las instrucciones de aplicación y de resolución son claras y específicas y las directrices para interpretarlas son excelentes. Por otro lado, la documentación completa es costosa de consultar debido a su extensión y estructura, y en ella no se encuentran evidencias sobre la eficacia de los programas propuestos para la mejora de aquellos casos con puntuaciones excesivamente bajas. En cambio, sí recomienda el uso de acomodaciones en aquellos casos en que sea necesario y siempre con la supervisión de un psicólogo o psicóloga, y ofrece recomendaciones para aquellos casos con problemas lingüísticos.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Batería de Actividades mentales Diferenciales y Generales Nivel M - renovado

1.2. Nombre del test en su versión original:

-

1.3. Autor/es del test original:

Carlos Yuste, David Yuste, Rosario Martínez Arias y 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

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:

La Batería de Actividades Mentales Diferenciales y Generales, nivel M renovado permite medir habilidades mentales básicas y generales así como la inteligencia general en los alumnos que se encuentran cursando la educación secundaria obligatoria (ESO). Su objetivo principal es discriminar habilidades mentales al tiempo que se obtienen estimaciones globales de actividades más generales, sin embargo también se usa para la selección profesional en niveles administrativos.

El test está formado por 6 escalas generales y 3 complementarias.

Escalas generales:

1. Analogías verbales Rv: Prueba específica de razonamiento y comprensión verbal.
2. Series numéricas Rn: Prueba específica de razonamiento serial numérico o habilidad para determinar regularidades lógicas en secuencias numéricas.
3. Matrices lógicas Re: Prueba de razonamiento inductivo, menos automatizado y menos trabajado culturalmente. Representan una modalidad del contenido que no se utiliza explícitamente en los programas educativos para enseñar ninguna asignatura ni impartir conocimientos. Se utilizan figuras geométricas donde éstas se relacionan lógicamente.
4. Completar oraciones Vv: Prueba de inteligencia verbal que requiere reconocimiento de vocabulario y situaciones sobre las que se deben tener conocimientos previos.
5. Problemas numéricos Nn: Prueba de rapidez de cálculo y razonamiento numérico que requiere de las aplicaciones de operaciones numéricas en problemas numéricos-verbales. No se pretende medir los conocimientos de matemáticas sino la correcta automatización de las correctas operaciones matemáticas básicas.
6. Encajar figuras Ge: Prueba que mide la habilidad para girar figuras geométricas, manteniendo sus relaciones de tamaño, distancia y posición relativas y comprobando la adecuación de figuras con la superficie de la que ha recortado.

Escalas complementarias:

1. Memoria de relato oral Ma: Prueba que mide la operación de recordar a corto plazo datos de un relato verbal, datos la mayor parte relacionados, dentro del sentido global de la lectura que se ha escuchado. Es una memoria verbal y numérica por el tipo de contenidos a recordar frente a una posible memoria que llamaríamos visoespecial.
2. Memoria visual ortográfica Mv: Prueba que mide la retención en la memoria a largo plazo de la correcta escritura de las palabras que aunque fonéticamente de pronuncien igual se escriben de distinta manera.
3. Discriminación de diferencias De: Prueba que mide la rapidez perceptiva, rapidez en operaciones simples de comparación de detalles entre figuras.

El Factor de Razonamiento, RR: suma de Rv, Rn y Re. El Factor Numérico, NN: suma de Rn y Nn El Factor Verbal, Vv: suma de Rv y Vv El Factor Visoespecial, EE: suma de Re y Ge. La Inteligencia General es la suma de las 6 pruebas básicas y a partir de él se puede estimar el coeficiente intelectual

(CI). Por otra parte es mide también la eficacia y la rapidez obtenida en las respuestas.

1.12. Áreas de aplicación:

Psicología Educativa

1.13. Formato de los ítems:

Elección múltiple

1.14. Número de ítems:

Seis pruebas de habilidades mentales básicas, de 32 ítems cada una, que conforman la inteligencia general (192 ítems). Tres pruebas complementarias adicionales de, igualmente, 32 ítems cada una.

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:

Estudiantes de Enseñanza Secundaria Obligatoria de España

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

Existen las formas en papel y lápiz y informatizada (on line) tanto para la corrección como para responder el test

1.19. Procedimiento de corrección:

Manual o automatizada por ordenador. También es posible mediante lectura óptica (facilitado por la empresa suministradora)

1.20. Puntuaciones:

La obtención de las puntuaciones globales de las distintas escalas generales (exceptuando la Prueba Numérica) y las 3 escalas complementarias tienen como fórmula de corrección la realización del sumatorio del número de aciertos, la cual cosa nos proporcionará la Puntuación Directa (PD). En cambio, en la Prueba Numérica, para obtener la Puntuación Directa, a los aciertos se le restan la mitad de los errores. Así pues, la puntuación máxima en cada una de las nueve escalas es de 32.

Por otro lado, los factores de segundo nivel se obtienen de la siguiente forma:

- Razonamiento lógico (RR): $R_v + R_n + R_e$. Puntuación directa máxima de 96.
- Verbal (VV): $R_v + V_v$. Puntuación directa máxima de 64.
- Numérico (NN): $R_n + N_n$. Puntuación directa máxima de 64.
- Viso-espacial (EE): $R_e + G_e$. Puntuación directa máxima de 64.

Finalmente, el factor de tercer orden, Inteligencia General se obtiene la puntuación del siguiente modo:

- Inteligencia General (IG): $R_v + R_n + R_e + V_v + N_n + G_e$. Puntuación directa máxima de 192.

La rapidez, RA, se contabiliza sumando las respuestas emitidas (EM) y omitidas (OM) en las seis pruebas

básicas. Se considera emitidas las marcas múltiples y como omitidas las preguntas saltadas o no respondidas. A partir de la última pregunta respondida no se consideran como omisión las que Aun no se han respondido. Se calcula a partir de la siguiente fórmula $RA=EM + OM$, siendo EM el número de respuestas emitidas en las seis pruebas básicas y OM el de las respuestas omitidas. La puntuación máxima posible es de 192.

Por otra parte, para obtener la puntuación de eficacia, EF, se obtienen dividiendo el número de aciertos o PD en IG, entre la suma de respuestas emitidas en las seis pruebas específicas (EM) y el número de doble marcas (DM). El resultado de esta división se multiplica por 100 y obtenemos el porcentaje de eficacia. La puntuación directa máxima posible es de 100.

1.21. Escalas utilizadas:

Centiles, eneatispos y CI de desviación

1.22. Posibilidad de obtener informes automatizados:

Sí: 6 tipos de informes automatizados:

Informe Colectivo. Resultados baremados y estadísticos de grupo: resultados en PD y Puntuación Centil en puntuaciones globales, las 6 pruebas básicas y las complementarias. Útil para el psicólogo y profesor tutor del grupo o sección para tener en un solo informe los principales estadísticos para comparar sus resultados con la muestra de referencia del baremo.

Perfil individual: se obtiene a partir de la puntuación percentil obtenida por el alumno, de manera gráfica. La PC del alumno se compara con los PC de los baremos en cada una de las pruebas.

Informe individual descriptivo: se muestra una valoración cualitativa textual en el rendimiento de cada prueba, a partir de las PC obtenidas en cada una de las variables baremadas.

Informe Intergrupos comparativo de medias, entre grupos o secciones de un curso: se compara la actividad mental en torno a la media aritmética de sus resultados, de grupos o secciones de un mismo curso en un centro determinado. La comparación es doble: primero se compara la media de cada grupo con la del baremo correspondiente, y segundo se compraran los diversos grupos/secciones entre sí. Útil para el psicólogo orientador y profesores para encontrar los puntos débiles de cada grupo y actuar en consecuencia.

Informe Intragrupo de distribución por niveles, de una sección o grupo: los PC se distribuyen en 7 niveles: muy alto, alto, medio-alto, medio, medio-bajo, bajo, muy bajo. También se compara la media obtenida por el grupo en cada factor, con la media del baremo. Útil para el psicólogo orientador y los profesores. Útil para saber con qué ritmo de enseñanza seguir en función del grupo y la asignatura impartida.

Informe sobre el intervalo de los verdaderos valores de IG y CI: para comprobar los límites superior e inferior entre los que se encuentran las puntuaciones percentiles "verdaderas" en cada alumno.

1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: 1º de la ESO: Pruebas básicas (67 minutos) y pruebas complementarias (24 minutos). Total: 91 minutos.

2º de la ESO: Pruebas básicas (64 minutos) y pruebas complementarias (22 minutos). Total: 86 minutos.

3º de la ESO: Pruebas básicas (58 minutos) y pruebas complementarias (20 minutos). Total: 78 minutos.

4º de la ESO y adultos: Pruebas básicas (54 minutos) y pruebas complementarias (18 minutos). Total: 72 minutos.

En cuanto a la administración Online varía mucho ya que depende de las características individuales del alumno. Normalmente, completan las 6 pruebas básicas en unos 50 minutos.

1.24. Documentación aportada por el editor:

Manual, información técnica complementaria on-line, cuaderno de ítems-elementos y cuaderno de respuestas.

1.25. Precio de un juego completo de la prueba:

Versión papel y lápiz: 55,06 € (manual, libro de elementos y 25 hojas de respuesta).

Licencia on-line: 48 € (60 usos, los 30 siguientes a 19,50 €, o los siguientes 300 usos a 149€) (20/12/2019)

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

4,78 € (1 unidad) (20/12/2019)

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

9,18 (25 unidades) (20/12/2019)

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

60 correcciones on line por 48€; si se entregan físicamente en papel, 6€ por alumno, y un incremento de 3€, por alumno (9€ en total), si se desea la entrega de informes en formato físico (20/12/2019)

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 - Excelente	4
2.3. Fundamentación teórica	Excelente	5
2.4. Adaptación del test	-	-
2.5. Desarrollo de los ítems del test	Excelente	5
2.6. Calidad de las instrucciones para el participante	Excelente	5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Excelente	5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Excelente	5
2.10. Datos sobre el análisis de los ítems	Buenos	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 y correlación con otras variables	
2.11.2.1.2. Tamaño de las muestras	Dos estudios con muestras de diferentes tamaños	4
2.11.2.1.3. Procedimiento de selección de las muestras	No se aporta información	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	Adecuada	3
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	-	-
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	Excelente	5

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	Notas en matemáticas, lengua, inglés y sociales al finalizar el curso	
2.11.2.2.2. Calidad de los criterios empleados	Buena	4
2.11.2.2.3. Relación temporal entre test y criterio	Predictiva	
2.11.2.2.4. Tamaño de las muestras	Un estudio con muestra grande (N > 200)	3
2.11.2.2.5. Procedimiento de selección de las muestras	No se aporta información	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Adecuado	3

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Adecuado	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	<p>hay algunas poblaciones y/o grupos minoritarios los cuales se tendrán que aplicar ciertas restricciones:</p> <ul style="list-style-type: none"> • Población cuya lengua materna no es el español: siempre que hable y entienda el español podrá hacer las pruebas Rn, Nn, Re y Ge pero no Rv, Vv y Ma. • Poblaciones inmigrantes: evitar pruebas visoespaciales al menos en sus primeros años. • Con las personas con diversidades funcionales se deberá estudiar cómo podría afectar a los resultados, y se podrían hacer las adecuaciones necesarias para su correcta realización a cargo de psicólogo/a que esté administrando el test. 	

2.11.5. Comentarios generales sobre evidencias de validez:

La documentación aporta información relativa a la validez desde las evidencias de contenido, de proceso de respuesta, de relación con otras variables (criterial predictiva y relación con sexo) y de estructura interna.

Respecto al contenido, aportan un excelente soporte teórico para definir y justificar el dominio que comprende la inteligencia, basándose en tres códigos/ámbitos, principalmente: verbal, numérico, visoespacial, en los que se da un peso importante al razonamiento inductivo y que además presentan una evolución continua a lo largo del periodo escolar

En el caso de los procesos de respuesta, se detallan muy bien las operaciones y procedimientos seguidos en la resolución de los distintos ítems, y además, se hace una clara distinción de las pruebas que pertenecen a cada factor de Inteligencia y los criterios utilizados para el ordenamiento teórico de los niveles de dificultad. Se establece también una distinción clara y precisa de las diferentes habilidades requeridas para cada prueba. También se aplica un análisis empírico de ítems muy preciso que invita a hacer una revisión de los mismos ya que, aunque en general los índices de dificultad son correctos, se producen algunas incongruencias y/o ambigüedades y por ello se recomienda también una revisión de los mismos por parte de personas expertas externas al equipo que ha desarrollado el test.

Por otro lado, también se describen cuatro estilos de respuesta basados en las escalas de Rapidez (RA) y Eficacia (EF) (Rapidez y Eficacia, Rapidez y Ineficacia, Lentitud y Eficacia, Lentitud y Ineficacia), y el estado del alumno al realizar la prueba en una única pregunta al finalizar el cuestionario (Problemas).

En el análisis de la estructura interna, se desarrollan únicamente resultados relativos a análisis exploratorios y en concreto, análisis en componentes principales. Dicho análisis presenta suficiente apoyo empírico para la consideración de valorar un único factor (inteligencia general) además de apoyar la correcta justificación de los 3 dominios en los que se divide esta prueba (verbal, numérico y visoespacial).

En cuanto a la relación de las puntuaciones del test con otras variables externas, se proporcionan tres tipos de evidencias: una de ellas se basa en establecer la igualdad de puntuaciones en función del sexo. Los análisis son excelentes en cuanto a muestra y comparaciones, y los resultados muestran diferencias estadísticamente significativas en algunas subescalas, algo que puede atribuirse a la elevado tamaño de la muestra. Aun así, la magnitud de estas diferencias es muy baja y no resulta relevante para los usos previstos del test.

Se facilita también la relación de las puntuaciones con criterios relevantes (rendimiento académico en lengua, matemáticas, inglés y sociales) como evidencia de validez predictiva. Las correlaciones, todas estadísticamente significativas, en general son adecuadas, aunque sería recomendable utilizar tests cuya medida sea el rendimiento escolar controlando el efecto atribuible a las variaciones curriculares entre centros.

Finalmente se ofrece una evidencia de validez convergente usando una forma casi paralela de BADyG/E3-r, el IGF/E, que, a pesar de ser similar y de haber sido extraída del mismo banco de ítems, solamente se compone de 70 ítems, número muy dispar respecto a los 192 ítems de BADyG/E3-r, por lo que no lo podemos considerar como una forma "casi paralela". Respecto a la validez discriminante, los autores muestran su disconformidad con su utilidad en general, aunque sería necesaria circunscribirlo a la medición que se propone.

Finalmente se hace mención y se incorporan diferentes programas de intervención a aplicar en aquellas personas con resultados significativamente bajos, aunque no se muestran evidencias empíricas de su eficacia.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de consistencia interna para cada una de las escalas y para diferentes grupos y error típico de medida en cada uno de ellos	

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	Alfa de Cronbach y dos mitades corregido	
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:

En líneas generales, se reportan resultados muy positivos de fiabilidad a partir de índices apropiados de consistencia interna, método de las dos mitades y alfa de Cronbach, pese a que este último se aplica a ítems con diferente grado de dificultad y por tanto no homogéneos, en todos los rangos de edad que abarca el test y con todas las subescalas. Se obtienen valores superiores a ,94 en el factor general, y superiores a ,90 en los factores de segundo orden, y superiores a ,75 en los de primer orden. También se aportan los datos de los errores típicos (Dmáx) y de los errores típicos de medida que se muestran lo suficientemente bajos como para considerar que las puntuaciones son precisas.

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 grandes	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	
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:

La muestra empleada es heterogénea y aporta una gran credibilidad a los resultados ya que esta, está formada por un gran número de participantes y nos permite inferir con más precisión a la hora de extrapolar las conclusiones. Los procedimientos para la interpretación de las puntuaciones del test son adecuados, teniendo en cuenta los objetivos de medición de la prueba. Se emplean muestras amplias y representativas para la elaboración de los baremos, y se describen de forma adecuada los estadísticos utilizados.

Para ello, se proporcionan cuatro baremos, muy actuales y en función del curso académico (de 1º a 4º de ESO). Las muestras de estos baremos son grandes por lo que se obtiene una alta precisión en los estadísticos calculados y se consigue la representatividad necesaria de la población objetivo. A este respecto, las variables controladas (tipo de escuela y comunidad autónoma) presentan algunas carencias: sobrerrepresentación de las escuelas privadas en la muestras de 2º, 3º y 4º de ESO, y ausencia de casos de varias comunidades autónomas en los diferentes niveles. Respecto a la primera, la diferencia en proporción de alumnos de escuela pública respecto a la población diana es del 20% en 4º de ESO y de más del 10% en 2º y 3º. Ello implica el tener cierta precaución a la hora de interpretar las puntuaciones con estos baremos, que son perfectamente útiles pese a esta desviación; respecto a la segunda carencia se considera que el constructo evaluado no presenta grandes diferencias en nivel ni en configuración entre las personas de las diferentes comunidades autónomas, por lo que es plausible aplicar los baremos a una comunidad aunque no esté representada en la muestra normativa.

Asimismo, se proponen claves para la correcta interpretación de los resultados, así como tipos de informes que se pueden proporcionar a el/los receptor/es. Respecto al establecimiento de puntos corte, encontramos que en el Informe Intragrupo de distribución por niveles, de una sección o grupo, clasifica a los estudiantes en siete niveles normativos para ver cuántos alumnos se sitúan en cada uno de ellos. Cada nivel tiene establecido un punto de corte en base al percentil: Muy alto (94), Alto (75), Medio-Alto (61), Medio (31), Medio-Bajo (16), Bajo (7), Muy Bajo (1). Los autores no explican el proceso de obtención de estos puntos ni qué criterios usan.

Por último, se presenta el error de medida de cada prueba, facilitando así el cálculo del intervalo en que se encontrará la puntuación verdadera. De esta forma, también se proporciona el error máximo con una confianza del 90%, y con una confianza del 95%.