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 la

Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas 3º y 4º E. Primaria

PAIB-2 Renovado

Prueba valorada en la ronda de evaluaciones del año 2018

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas (PAIB-2 Renovado): 3º y 4º de Educación Primaria
Autor	José Luis Ramos Sánchez, José Luis Galve Manzano y Rosario Martínez Arias
Autor de la adaptación española	---
Variable(s)	Competencias instrumentales básicas en lenguaje y matemáticas
Áreas de aplicación	Psicología educativa
Soporte	Papel y lápiz, informatizada

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Adecuada-Buena	3,5
Fundamentación teórica	Adecuada	3
Adaptación	---	---
Análisis de ítems	Adecuada-Buena	3,5
Validez: contenido	Adecuada, pero con algunas carencias	2,5
Validez: relación con otras variables	Adecuada	3
Validez: estructura interna	No se aporta información	---
Validez: análisis del DIF	No se aporta información	---
Fiabilidad: equivalencia	No se aporta información	---
Fiabilidad: consistencia interna	Excelente	5
Fiabilidad: estabilidad	No se aporta información	---
Fiabilidad: TRI	No se aporta información	---
Fiabilidad inter-jueces	No se aporta información	---
Baremos e interpretación de puntuaciones	Buena	4

Comentarios generales

La Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas (PAIB-2-Renovado) evalúa los aspectos instrumentales básicos en la lectura y en la escritura, relacionadas con el currículo escolar vigente y desde la perspectiva de la psicología cognitiva, en estudiantes de 3º y 4º de Educación de Educación Primaria. Las competencias evaluadas son relevantes y las tareas consideradas resultan pertinentes y representativas de los objetivos curriculares en estos niveles. Por tanto, las pruebas pueden resultar de utilidad en el contexto educativo, tanto para hacer una valoración inicial a principio de curso, como para hacer una valoración del aprendizaje al finalizar este. El ámbito de aplicación del cuestionario es eminentemente educativo.

Entre los puntos fuertes de la prueba cabe destacar:

- El material resulta atractivo para estudiantes de 3º y 4º de Primaria, es de fácil manejo y las instrucciones de aplicación y corrección son claras. El manual es completo, con amplia justificación teórica y estadística, siendo sus contenidos accesibles para profesores y orientadores.
- Destaca el procedimiento seguido para la construcción de las pruebas, con tres fases que abordan pilotajes cualitativos y cuantitativos como base para la elaboración del instrumento final. Se justifica de forma teórica el dominio de representación de los diferentes ítems.
- La selección de la muestra es incidental, pero se ha procurado que exista variabilidad en cuanto a representación geográfica, tipología de los centros (público o privado/concertado; centros urbanos o rurales) y distribución por sexo. El uso de cuotas pone cierto control a los sesgos, aunque no permita precisar cuán representativa es la muestra.
- Los resultados proporcionados como evidencias de fiabilidad de las puntuaciones (coeficientes alfa de Cronbach) son excelentes para las puntuaciones globales en Lenguaje y Matemáticas. Para las subescalas se consideran buenos, siendo algo más bajos pero adecuados para las pruebas de Ortografía y Medida (en ambos niveles educativos) y en Cálculo (para 3º de primaria). Además, se presentan IC95%, índice de discriminación de los ítems y valor de alfa eliminando cada uno de los ítems.
- Se aportan evidencias de validez basadas en relaciones con criterios externos. Más concretamente, en las relaciones con el criterio subjetivo de dos profesores (sobre el rendimiento de los estudiantes en lenguaje y matemáticas) y con la batería BADYG-R (Nivel E-2). En relación con el criterio subjetivo de los profesores, los coeficientes de validez llegan a ser excelentes para las puntuaciones globales en 4º de Primaria (Matemáticas y Lenguaje), aunque sería recomendable ampliar la muestra de 3º de Primaria (n=88).
- Se valora muy positivamente que se permita realizar aplicación y corrección tanto manual como informatizada, estando los baremos disponibles en centiles, puntuaciones T y decatipos. Además, el formato de informe de resultados en función de los baremos propuestos supone uno de los puntos fuertes de la herramienta, puesto que permite obtener una evaluación normativa tanto cuantitativa (percentiles, puntuaciones T, Decatipos) como cualitativa (escala ordinal de siete grados entre Muy Bajo y Muy Alto).

En términos generales, las propiedades psicométricas de la prueba se consideran buenas y legitiman su uso en el contexto educativo, si bien en futuras ediciones sería deseable tener en cuenta los siguientes aspectos:

- Sería recomendable incluir referencias más actualizadas y comprobar que exista una correspondencia exacta entre la lista de referencias y las citas en el texto del manual. Aunque resulta comprensible que exista una gran concentración de citas de trabajos de los autores de la prueba (volcados en la investigación en lenguaje y matemáticas en este campo), también resultaría deseable incluir alguna referencia a trabajos de otros grupos y otros países.

- Dado que se trata de versiones revisadas de las pruebas PAIB publicadas en 2009, resulta necesario especificarlo en el manual, así como proporcionar detalles acerca de qué aspectos se han renovado, eliminado o incluido. Del mismo modo, convendría aclarar si se utiliza en esta versión la misma muestra que en las pruebas originales.
- En los manuales del PAIB-1, PAIB-2 y PAIB-3 coinciden una gran cantidad de páginas, por lo que parece recomendable aunar todas las pruebas en un mismo manual y especificar los apartados necesarios cuando sea necesario distinguir entre ellas.
- Resulta recomendable mejorar el rigor en el uso de términos metodológicos y añadir en el manual datos específicos acerca de los procedimientos que indican haberse aplicado (p. ej., evidencias de validez de contenido basadas en la consulta a expertos, modo de selección y participación del profesorado).
- Aunque el análisis de ítems arroja buenos resultados, sería interesante incluir índices de validez de contenido, pesos factoriales en la estructura latente y evidencias de validez criterial.
- Ya que varios de los ítems que forman parte de la versión final cuentan con índices inferiores a valores mínimos aceptables, parece necesario llevar a cabo una depuración del instrumento eliminando aquellos con capacidad discriminativa inadecuada,
- Sería interesante, además del alfa de Cronbach, aportar más estimaciones de la fiabilidad de las puntuaciones y tener en cuenta que los ítems de distintas escalas reciben distinta ponderación.
- Sería deseable incorporar evidencias de validez basadas en la estructura interna del PAIB (p. ej., análisis factorial confirmatorio, análisis del método de la varianza común). También, se sugiere el contraste de una red nomológica que incluya diferentes variables criterio desde una perspectiva del modelado de ecuaciones estructurales.
- Podrían considerarse evidencias de validez convergente más directas y objetivas, en relación con otras pruebas específicas, estandarizadas, de competencia curricular. De hecho, se podría introducir una estrategia de obtención de evidencias de validez de decisión respecto a una regla de oro acerca de las competencias de Lenguaje y Matemáticas, y llevar a cabo análisis de curvas ROC, sobre puntuaciones totales o sobre ítems.
- Sería interesante incluir más información sobre la validez discriminante de las puntuaciones (p. ej., en qué grado las puntuaciones en las subescalas de Lenguaje y Matemáticas están diferenciadas como se espera y de forma suficiente para que se puedan establecer perfiles de competencias, ya que se observa la presencia de un fuerte factor general).
- Existen erratas e imprecisiones que deben subsanarse en futuras ediciones.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

PAIB-2 Renovado. Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas. 3º y 4º de Educación Primaria.

1.2. Nombre del test en su versión original:

PAIB-2. Prueba de Aspectos Instrumentales Básicos en Lenguaje y Matemáticas (3º y 4º Educación Primaria)

1.3. Autor/es del test original:

José Luis Ramos Sánchez, Jose Luis Galve Manzano y Rosario Martinez Arias

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:

2009.

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

1.9. Fecha de la última revisión del test:

2017.

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

Rendimiento académico / competencia curricular.

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

El PAIB-2 renovado, dirigido a los niveles de 3º y 4º de Educación Primaria tiene como objetivo la evaluación de aspectos instrumentales básicos en lenguaje y matemáticas, relacionadas con el currículo escolar. Basado en la psicología cognitiva, trata de valorar tanto los procesos cognitivos puestos en marcha como sus productos. Está concebida como una prueba pedagógica que ayude al profesorado y profesionales de la orientación educativa a determinar el nivel de desarrollo alcanzado por el alumnado en lenguaje y matemáticas.

De forma específica, en cuanto al Lenguaje (3 subescalas), evalúa:

- Vocabulario: a partir de una tarea de reconocimiento de sinónimos y de otra que consiste en completar palabras.
- Ortografía: mediante la escritura de palabras de diversa dificultad ortográfica.
- Comprensión lectora: a partir de la lectura de un texto de 300 palabras sobre "los Hipopótamos" y la posterior respuesta (sin apoyo del texto) a cuestiones sobre el propio texto.

En cuanto a las competencias matemáticas (4 subescalas), el PAIB-2 evalúa:

- Numeración: la comprensión e interpretación de los números y su escritura, utilizando números naturales, fraccionarios y romanos.
- Cálculo: la realización de las operaciones de cálculo matemático básico (suma, resta, multiplicación y división), así como la aplicación de paréntesis y orden de prioridad de las operaciones.
- Medida: la competencia para resolver problemas relacionados con el manejo de los diferentes sistemas de medida (moneda, longitud, superficie, tiempo, masa, capacidad).
- Resolución de problemas: la competencia de resolver problemas que requieren del uso de las operaciones matemáticas básicas.

El PAIB-2 presenta una doble utilidad dentro del ámbito educativo, pudiendo ser empleado como evaluación inicial al principio de cada curso en Educación Primaria o evaluación final, otorgándole una naturaleza de prueba final de madurez con el ánimo de poder diseñar contenidos curriculares que permitan al alumnado mejorar el nivel de desarrollo del alumnado pertinente de cara al curso posterior o los meses que conforman el mismo y prevenir así potenciales dificultades de aprendizaje. Igualmente, el PAIB-2 presenta la posibilidad de ser aplicado en cualquier momento del curso de cara a valorar el nivel de aprendizaje del alumnado, incluidos aquellos con dificultades de aprendizaje o compensación educativa.

La prueba está diseñada para llevar a cabo una aplicación colectiva, aunque también es posible su aplicación de manera individual, requiriendo un total dos sesiones de 40 minutos cada una, con al menos media hora de descanso entre las sesiones de evaluación.

1.12. Áreas de aplicación:

Psicología educativa.

1.13. Formato de los ítems:

Respuesta construida y elección múltiple.

1.14. Número de ítems:

Lenguaje: 64 ítems en total, organizados en:

Vocabulario= 30 ítems, que se dividen en:

Sinónimos= 15 ítems
Palabras incompletas= 15 ítems

Ortografía= 16 ítems
Comprensión Lectora= 18 ítems

Matemáticas: 60 ítems en total, organizados en:

Numeración= 16 ítems
Cálculo= 16 ítems
Medida= 16 ítems
Resolución de problemas= 12 ítems

1.15. Soporte:

Papel y lápiz. Informatizado.

Corrección manual o por internet.

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

Nivel A.

En el manual se indica que las pruebas pueden ser aplicadas por el profesorado de Lenguaje y Matemáticas, de apoyo, psicólogos, psicopedagogos, pedagogos y orientadores, familiarizados con el enfoque cognitivo.

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

Alumnos de 3º y 4º de Primaria (8-10 años).

No se aporta información específica sobre su aplicación a poblaciones específicas o minoritarias.

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

Dispone de versión en lápiz y papel (con posibilidad de corrección informatizada a través de la web del editor) y versión informatizada (con aplicación online a través de la web del editor). No requiere hardware o software específico ya que se administra y puede corregirse en una plataforma de corrección on-line (paib.cepeonline.es).

Ha de tenerse en cuenta que existen también con el mismo nombre el PAIB-1 (Ed. Infantil, 1º y 2º de Educación Primaria) y el PAIB-3 (5º y 6º de Educación Primaria y 1º de ESO), siendo estas versiones renovadas de la prueba original publicada en 2009.

1.19. Procedimiento de corrección:

La corrección puede realizarse tanto de forma manual como de forma automatizada, mediante una plataforma de corrección on-line.

1.20. Puntuaciones:

En todos los casos los ítems están planteado de manera positiva, indicando un acierto una mayor puntúación en la variable evaluada. Las puntuaciones directas se obtienen mediante la suma de los aciertos obtenidos en cada uno de los apartados. Se calcula una puntuación directa total para cada uno de los bloques incluidos en las pruebas: lectura, escritura y matemáticas.

Resulta posible obtener un informe individual y colectivo, obteniendo puntuaciones en diversas escalas normalizadas y no normalizadas (centiles, T y decatipos), así como una interpretación cualitativa ordenada por rangos.

El manual ofrece información clara y detallada en cuanto a la otorgación de puntuaciones por acierto. Se desconoce, ya que no se especifica en el manual, la posibilidad de tener en cuenta potenciales aciertos debidos al azar.

Las puntuaciones pueden obtenerse de manera manual, en la versión papel y lápiz, o bien mediante la corrección automatizada del instrumento, estando, por tanto, este procedimiento completamente automatizado.

1.21. Escalas utilizadas:

Centiles, decatipos, puntuaciones T (M=50; DT=10) y estaninos.

1.22. Posibilidad de obtener informes automatizados:

Sí. El programa informatizado on-line proporciona los resultados en forma de perfil individual y registro grupal. El informe automatizado que genera el PAIB-2 consiste en varias tablas independientes que recogen las puntuaciones (centiles, puntuaciones T y decatipos) de las subescalas (vocabulario, ortografía, comprensión lectora, numeración, cálculo, medida, resolución de problemas) y globales de la prueba (global lengua, global matemáticas). Para cada prueba se describe brevemente la competencia evaluada y se clasifica el rendimiento en "Muy bajo", "Bajo", "Medio", "Alto" y "Muy alto", sin más detalle. Las gráficas de perfil son monocromas, claras, muy visuales y fácilmente interpretables, recordando el área evaluada por cada prueba e indicando gráficamente la posición de la persona con respecto a la población.

1.23. Tiempo estimado para la aplicación del test:

Tanto en aplicación individual como colectiva, el tiempo estimado para la aplicación es aproximadamente de 1 hora 20 minutos. Se recomiendan dos sesiones de 40 minutos cada una (con 30 minutos de descanso entre las sesiones de evaluación).

1.24. Documentación aportada por el editor:

Manual y cuadernos de trabajo del alumno.

Licencia para aplicación y corrección online.

1.25. Precio de un juego completo de la prueba:

41,10 euros (Manual).

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

2,33 euros cada cuaderno de trabajo.

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

--

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

49,95 euros (licencia on-line).

24,95 euros (renovación de 50 usos).

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Buena	4
2.2. Calidad de la documentación aportada	Adecuada	3
2.3. Fundamentación teórica	Adecuada	3
2.4. Adaptación del test	---	---
2.5. Desarrollo de los ítems del test	Adecuada	3
2.6. Calidad de las instrucciones para el participante	Buena	4
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Buena-Excelente	4,5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Adecuada, pero con algunas carencias	2
2.10. Datos sobre el análisis de los ítems	Adecuada-Buena	3,5

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	Adecuada	3
2.11.1.2. Consultas a expertos	Se ha consultado de manera informal a un pequeño número de expertos	2

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	Correlaciones con otro test	
2.11.2.1.2. Tamaño de las muestras	Dos estudios con muestras de tamaño moderado	4
2.11.2.1.3. Procedimiento de selección de las muestras	Incidental	
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	Adecuada	3
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	No se aporta	---
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	No se aporta	---
2.11.2.1.8. Resultados de las diferencias intergrupo	No se aporta	---

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	Correlación con valoración subjetiva del profesorado sobre el rendimiento escolar en lenguaje y matemáticas (escala de 1=nivel muy bajo a 5=nivel muy alto)	
2.11.2.2.2. Calidad de los criterios empleados	Adecuada, pero con algunas carencias	2
2.11.2.2.3. Relación temporal entre test y criterio	Concurrente	
2.11.2.2.4. Tamaño de las muestras	Un estudio con una muestra pequeña y otro con una muestra moderada	3,5
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Buena	4

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	---	---
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	No	

2.11.5. Comentarios generales sobre evidencias de validez:

En el manual se incluyen apartados de validez centrados en la relación con otro test (BADYG-R) y con un criterio externo (criterio subjetivo de dos profesores sobre el rendimiento de los estudiantes en lenguaje), validez de contenido y validez estructural. Aunque en algunos casos los autores utilizan la expresión "evidencias de validez", no utilizan una concepción de validez unitaria (cfr. AERA, APA y NCME, 2014).

Las evidencias de validez aportadas resultan en términos generales adecuadas, pero para futuras ediciones de la prueba se recomienda:

- Aportar evidencias específicas basadas en el contenido: número de expertos consultados, procedimiento de selección, procedimiento, análisis y resultados obtenidos sobre el grado de coincidencia, etc. También se recomienda revisar aquellos ítems con correlaciones ítem-total por debajo de .200.
- Proporcionar evidencias de validez basadas en la estructura interna adicionales: el manual incluye un breve apartado referido a "validez estructural" que se limita a presentar las correlaciones entre las distintas escalas, utilizándose esta información para justificar que existe una "habilidad subyacente" a todas las pruebas. Sin embargo, resulta fundamental aportar datos relativos a otros procedimientos que no sólo indiquen la relación entre las variables, sino su aportación a cada factor y su ajuste al modelo dimensional que subyace al instrumento. Sería recomendable la utilización de análisis factorial confirmatorio para justificar las puntuaciones proporcionadas por el instrumento en las escalas y subescalas. Además, los resultados descritos en el apartado "Validez estructural" se han obtenido con muestras de distinto tamaño a las descritas en la sección Participantes en el estudio definitivo (*N*=705). La razón de estas discrepancias debería señalarse explícitamente.
- Complementar las evidencias basadas en un criterio externo pues se emplea como criterio la valoración subjetiva de "al menos dos profesores" sobre el nivel de aprendizaje en el área de Lenguaje en una escala de 1 (no supera los objetivos previstos) a 4 (supera los objetivos previstos). Además de que se trata de un criterio subjetivo que puede estar sesgado, se desconocen las características y número de los respondientes, los criterios que utilizaron para realizar tal valoración y la fiabilidad interjueces. Los análisis se limitan a correlaciones (aunque se indica en el manual que también se realiza regresión lineal múltiple, los datos de esta no se encuentran en el mismo) entre las valoraciones subjetivas de los profesores y las diferentes subescalas y puntuaciones globales. Sería interesante utilizar procedimientos para la estimación de tales relaciones considerando el error de medida y utilizar una potencial "regla de oro" para posibilitar un análisis de validez de decisión de las puntuaciones obtenidas. Cabe subrayar también que los tamaños de las muestras son muy desiguales (88 alumnos evaluados en 3º EP frente a 432 en 4º EP).
- En cuanto a las evidencias de validez basadas en la relación con otros tests, se proporcionan las correlaciones con la Batería de Aptitudes Diferenciales y Generales (BADYG-R), en su nivel elemental (E-2: 3º y 4º de EP). Los valores obtenidos son adecuados y muestran que existe relación entre las dimensiones evaluadas en ambas pruebas, si bien para el nivel de 3º de EP no se proporcionan). Sin embargo, seria conveniente incluir una justificación sobre los motivos para elegir esta prueba. Para valorar la validez convergente y discriminante de las pruebas de forma más específica sería recomendable aplicar otras pruebas estandarizadas de rendimiento escolar.
- Se sugiere realizar análisis de posibles diferencias estadísticamente significativas entre grupos de interés y análisis del funcionamiento diferencial de los ítems.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Un único coeficiente de fiabilidad (para cada escala o subescala) y 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 y una de tamaño moderado	4,5
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa o KR-20	
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:

Las muestras utilizadas para analizar la fiabilidad de las puntuaciones se consideran adecuadas en cuanto a la heterogeneidad existente en términos de sexo, distribución geográfica y titularidad de los colegios a los que pertenecen los participantes. La muestra global (N=705) y de 4º de EP son grandes, la submuestra de 3º de primaria es de tamaño moderado (n=205).

Se informa de un coeficiente de fiabilidad, el coeficiente alfa de Cronbach, que indica consistencia interna, para cada escala, subescala y puntuación global en cada uno de los subgrupos poblacionales (3º y 4º e. primaria). Junto con los coeficientes alfa de Cronbach se presentan sus intervalos de confianzas, media y desviación típica, los índices de discriminación de los ítems (homogeneidad corregida) y el valor de alfa eliminando cada uno de los ítems. En futuras ediciones del test sería interesante utilizar otro tipo de estimadores que consideren tanto la información de cada una de las subpruebas como el número de elementos que conforman las mismas.

Los coeficientes son excelentes para las puntuaciones globales en Lenguaje y Matemáticas. Para las subescalas concretas, se encuentran valores buenos, salvo para las pruebas de Ortografía y Medida (en ambos niveles educativos) y en Cálculo (para 3º de primaria). En algunas de las subescalas se obtienen valores por debajo .80 e índices de discriminación inferiores a .20, cuya eliminación supondría una ganancia en la fiabilidad total de la escala. También sería interesante llevar a cabo análisis de fiabilidad con Teoría de Respuesta a los Ítems (TRI), por ejemplo, con el modelo de Rasch.

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	Moderado	3
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	Incidental	
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:

Las muestras empleadas se pueden considerar adecuadas en cuanto a tamaño y composición (muestreo por cuotas). El manual del PAIB-2 presenta dos conjuntos de baremos, distinguiendo entre 3º de EP (final de 3º, principio de 4º de EP) y 4º de EP (final de 4º, principio de 5º de EP). La estructura de baremos que presenta es la misma para ambos cursos:

- Tabla de puntuaciones directas y sus correspondientes percentiles para las subescalas de competencias en Lenguaje y en Matemáticas, y para las puntuaciones globales. La tabla también incluye la media y la desviación estándar.
- Tabla de puntuaciones directas y sus correspondientes puntuaciones T para las subescalas de Lengua y Matemáticas, más la media y la desviación estándar en escala T.
- Tabla de puntuaciones directas y sus correspondientes puntuaciones T para la escala global de Lengua y de Matemáticas, más la media y la desviación estándar en escala T.
- Tabla de puntuaciones directas y sus correspondintes puntuaciones eneatipo (estaninos) y decatipo, tanto para las subescalas de Lengua y Matemáticas como para las puntuaciones globales.

Los baremos permiten valorar la situación específica del alumno en cada una de las pruebas respecto a su grupo normativo, con una escala ordinal de 7 niveles (si bien esta clasificación parece basarse únicamente en el criterio estadístico de desviación típica): Muy bajo (P<10, T<30, DEC<=2), Bajo (P entre 11-24, T entre 30-39, DEC=3), Medio-Bajo (P entre 25-39, T entre 40-44, DEC=4), Medio (P entre 40-60, T entre 45-55, DEC=5), Medio-alto (P entre 61-75, T entre 56-60, DEC=6), Alto (P entre 76-89, T entre 61-70, DEC=7) y Muy alto (P>89, T>70, DEC>7).

Sería recomendable realizar estudios de diferencias intergrupos para comprobar si resulta necesario contar con baremos específicos para diferentes muestras dentro de cada estrato educativo (p. ej., en función del sexo u otras variables relevantes).

Se valora muy positivamente el proporcionar la posibilidad de realizar la aplicación y corrección de la prueba tanto de forma manual como de forma informatizada, considerándose como punto fuerte la inclusión de los baremos en el manual.