



Consejo General de
Colegios Oficiales
de Psicólogos

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 LAS ESCALAS

MPR

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Características	Descripción
Nombre del test	Merrill-Palmer-R. Escalas de desarrollo Merrill-Palmer revisadas
Autor	G.H. Roid, J. Sampers, G. Anderson, J. Erickson y P. Post
Autor de la adaptación española	Fernando Sánchez, Pablo Santamaría, Irene Fernández y David Arribas
Editor de la adaptación española	TEA ediciones
Fecha de la última revisión del test en su adaptación española	2011
Constructo evaluado	Desarrollo (cognitivo, lenguaje, motor, socio-emocional y conducta adaptativa)
Área de aplicación	Escolar, Clínica, Neuropsicología, Servicios sociales
Soporte	Administración oral. Manipulativo. Papel y lápiz

Valoración general

Características	Valoración	Puntuación
Calidad de los Materiales y documentación	Excelente	5
Fundamentación teórica	Buena	4
Adaptación Española	Excelente	5
Análisis de los ítems	Buena	4
Validez de contenido	Buena	4
Validez de constructo	Excelente	4.5
Análisis del sesgo		
Validez predictiva		
Fiabilidad: equivalencia		
Fiabilidad: consistencia interna	Excelente	4.5
Fiabilidad: estabilidad	Excelente	4.5
Baremos	Buena	4

Comentarios generales

La detección temprana de problemas y deficiencias en el desarrollo es fundamental para minimizar sus efectos en la vida posterior. Es importante además poder establecer comparaciones intra e interindividuales. Las primeras para detectar lagunas o inconsistencias y las segundas para detectar retrasos en el desarrollo. Los expertos señalan además la importancia de que las puntuaciones proporcionadas por los instrumentos de evaluación no tengan efectos suelo ni techo, de modo que puedan aplicarse a sujetos de grupos extremos en todo el continuo de habilidad y con adecuada precisión (Alfonso y Flanagan, 2008). Considerando lo anterior, la Escala MP-R parece muy adecuada para la exploración del nivel de desarrollo intelectual y de otros dominios, en un amplio rango de edad.

La labor de adaptación y actualización de la antigua escala (editada por primera vez en 1931), realizada por Roid y Sampers, ha sido excelente, así como la adaptación realizada por TEA. Además, en la actualidad, no existe en España ninguna otra prueba con las mismas características. La adaptación a la población española sólo tiene unos años de diferencia respecto a la edición original norteamericana de 2004.

Desde el ámbito profesional, los cinco dominios que evalúa son los especificados por la legislación federal de los Estados Unidos de América para el tratamiento de discapacidades infantiles. Aunque en España la reglamentación se encuentra dispersa, estos dominios también son fundamentales para los Equipos de Orientación Educativa y Psicopedagógica y de Atención Temprana.

En general, podemos considerar que la Escala goza de buenas propiedades psicométricas, que es fácil de aplicar y que el material parece atractivo para el niño. Como **puntos fuertes** se pueden indicar los siguientes:

1. Desde el punto de vista teórico, intenta enlazar con las modernas teorías de la inteligencia (Modelo Cattell-Horn-Carroll), contemplando aspectos de la inteligencia fluida, cristalizada, memoria y procesamiento viso-espacial.
2. En cuanto al diseño del test, cabe destacar: a) que ofrezca a la vez un Índice general de desarrollo cognitivo y la evaluación de 5 dominios (desarrollo cognitivo, lenguaje y comunicación, desarrollo motor, desarrollo socio-emocional y conducta adaptativa) que cubren la mayor parte de los requisitos de diagnóstico de discapacidades o retrasos en la temprana infancia, b) Su extenso intervalo de edades de aplicación (desde un mes a 6 años y medio) que abarca todo el período pre-escolar, y c) la escasa dependencia del lenguaje oral, lo que la hace adecuada para niños y niñas con dificultades lingüísticas:

trastornos del espectro autista, sorderas y deficiencias auditivas y otros retrasos en el desarrollo. Además, es especialmente importante señalar que la mayoría de las tareas de la prueba tienen un carácter lúdico, utilizándose juguetes y materiales atractivos para los niños de estas edades.

3. Presenta muy buenas propiedades psicométricas. Son elevados los indicadores de fiabilidad, entendida como consistencia interna y como estabilidad test-retest, y el Manual aporta variadas evidencias de validez que justifican los usos más comunes de las puntuaciones. La utilización de la Teoría de la Respuesta al Ítem permite logros no alcanzables con la Teoría Clásica, como son la obtención de la precisión de cada puntuación (en vez de solo la precisión del test), la obtención de las puntuaciones de desarrollo, y determinar las respuestas típicas que se pueden dar en cada nivel, gracias a que la dificultad de los ítems están en la misma escala que la habilidad. Otro avance psicométrico reseñable es que el Manual preste atención a las adaptaciones requeridas por los grupos especiales.
4. El tratamiento de todo lo relativo a la interpretación de las puntuaciones ha recibido mucha atención. Se ofrecen baremos referidos a las normas en términos de percentiles y escalas típicas de CI, con intervalos de un mes durante el primer año y de dos o tres meses en períodos posteriores. La evaluación tiene en cuenta diversos aspectos cualitativos, como el comportamiento del evaluado durante la evaluación. Se ofrecen además normas de interpretación referidas a criterios en términos de Puntuaciones de desarrollo y Edades equivalentes obtenidas mediante la Teoría de la Respuesta al Ítem.
5. Los materiales y documentación son muy buenos. Los estímulos y los materiales para el registro de las puntuaciones son excelentes y facilitan mucho la labor de los examinadores. Los cuadernillos facilitan la aplicación de las normas de corrección, que son muy claras.

Entre los **puntos débiles** y sugerencias, que pudieran atenderse en futuras ediciones, se podrían señalar los siguientes:

1. En cuanto a las bases teóricas, aunque se ha hecho el esfuerzo de justificar la prueba dentro del modelo CHC de Inteligencia y de los procesos subyacentes a las diversas tareas, no queda del todo claro en qué medida los resultados obtenidos soportan o no la teoría.
2. En el desarrollo de la versión americana se hicieron estudios sobre el sesgo de los ítems en relación al sexo y etnia. Se obtuvo que algún ítem mostraba funcionamiento diferencial. Convendría explorar la existencia de sesgo en los ítems de la versión española. En este mismo sentido, pudiera tener sentido explorar más a fondo las

posibles diferencias entre sexos y considerar la conveniencia de generar baremos diferentes para niños y niñas.

3. A pesar de lo extenso y completo que es el manual (522 páginas), se echan en falta ciertas informaciones: una descripción más completa del procedimiento de selección de las muestras, los detalles del procedimiento seguido para determinar el ajuste al modelo de Rasch, y detalles, resultados y una interpretación más completa de los análisis factoriales.
4. La aplicación total de la prueba requiere un tiempo cercano a una hora. La batería cognitiva, entre 30 y 40 minutos. Podría ser interesante disponer de una versión breve, formada por una selección de tareas, que pudiese servir como prueba de cribaje para la detección temprana de problemas en las áreas de desarrollo de más interés.
5. Algunos materiales son frágiles o de pequeño tamaño. Sería recomendable que fuera posible su reposición en caso de deterioro o pérdida. Adicionalmente, cabe sugerir que los editores incluyeran un DVD con información audiovisual con ejemplos de aplicación de las pruebas que planteen más dudas a los profesionales.

En resumen, la escala de desarrollo MP-R tiene un considerable interés para la evaluación de niños menores de 6 años y medio. La descripción que el Manual ofrece de los estudios realizados sobre el test original y en la adaptación española justifica su uso para la evaluación de los cinco dominios de desarrollo que la escala mide.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test

Merrill-Palmer-R. Escalas de desarrollo Merrill-Palmer revisadas

1.2. Nombre del test en su versión original

Merrill-Palmer- Revised. Scales of Development.

1.3. Autor

G.H. Roid, J. Sampers, G. Anderson, J. Erickson y P. Post.

1.4. Autor de la adaptación española

Fernando Sánchez Sánchez, Pablo Santamaría Fernández, Irene Fernández Pinto y David Arribas Águila

1.5. Editor del test en su versión original

Stoelting Company, Wood Dale, Illinois

1.6. Editor de la adaptación española

TEA Ediciones, S.A. Madrid

1.7. Fecha de publicación del test original

2004

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

2011

1.9. Fecha de la última revisión del test en su adaptación española

2011

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

- Inteligencia
- Psicomotricidad
- Escalas de Desarrollo
- Escalas Clínicas

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

La escala *Merrill-Palmer-R*, publicada en 2004, es una actualización de una escala clásica de desarrollo del mismo nombre publicada en 1931. Permite evaluar cinco dominios de desarrollo entre las edades de 1 mes y 78 meses (6 años y medio):

- *Desarrollo cognitivo* mediante tareas que permiten poner de relieve el razonamiento verbal y no verbal, memoria, aspectos viso-motores y rapidez de procesamiento.
- *Desarrollo del lenguaje y comunicación* por medio de la evaluación del lenguaje receptivo y expresivo, evaluados por el examinador y los padres o cuidadores del niño.
- *Desarrollo motor*. Evaluación de la motricidad fina y gruesa.

- *Conducta socio-emocional*, evaluada mediante dos cuestionarios (uno para aspectos evolutivos y otro para los no evolutivos o temperamento), junto con datos clínicos proporcionados por los padres.
- *Conducta de autocuidado y adaptativa*, evaluada por padres o cuidadores.

Los cinco dominios son evaluados por medio de actividades manipulativas e instrucciones orales, así como con escalas de valoración cumplimentadas por padres y examinadores. Puede obtenerse una evaluación global (Índice de Desarrollo Global) del funcionamiento intelectual y puntuaciones separadas para cada dominio.

La batería de Desarrollo cognitivo está formada por tres escalas principales, tres complementarias y dos escalas infantiles especiales (para niños menores de 12 meses). Las tres principales son: Cognición, Motricidad fina y Lenguaje receptivo. Las tres complementarias: Memoria, Velocidad de procesamiento y Coordinación viso-motora. Las dos escalas infantiles especiales son Lenguaje infantil y Memoria infantil. Todas las escalas cognitivas, excepto el Lenguaje infantil, forman parte del Índice de desarrollo global.

El Lenguaje Expresivo se evalúa mediante dos procedimientos, el lenguaje expresivo-padres y el lenguaje expresivo-examinador. Se complementa con Lenguaje receptivo para tener un índice total de lenguaje.

La escala de Motricidad gruesa evalúa varios aspectos del desarrollo motor mediante 23 pruebas breves con formato tipo juego.

El dominio Socio-emocional puede evaluarse mediante 4 instrumentos:

- Escala socioemocional cumplimentada por padres o cuidadores
- Estilo de temperamento, cumplimentada por padres o cuidadores
- Comportamiento durante la evaluación, evaluado por el examinador
- Indicadores de problemas socioemocionales o conductuales, que es una entrevista estructurada que se puede aplicar si el examinador detecta indicios de problemas.

La Conducta adaptativa y de autocuidado se evalúa mediante el examen de diferentes aspectos relacionados con el desarrollo de la autonomía e independencia: alimentarse, vestirse, asearse, conciencia de peligros, etc. La información se recoge por medio de un cuestionario cumplimentado por los padres.

Esta prueba se considera especialmente indicada para evaluar la posible presencia de trastornos del desarrollo en niños, escolarizados o no, en Educación Infantil, así como para evaluar el progreso en su desarrollo en las distintas áreas mediante la elaboración de un perfil de desarrollo. Debido a que en su elaboración se ha utilizado el modelo de Rasch, es especialmente recomendable su uso para analizar el progreso individual de cada niño entre dos aplicaciones distanciadas en el tiempo, permitiendo ofrecer una información útil tanto a padres como a profesores de los avances en las distintas áreas de desarrollo. Al ser una revisión de las Escalas Merrill-Palmer, la presencia de elementos verbales en las distintas escalas (excepto en las de Lenguaje) es muy escasa, de modo que es posible su uso en niños con escaso conocimiento de la lengua o con problemas de comunicación.

1.12. Áreas de aplicación

- Psicología clínica
- Psicología educativa
- Neuropsicología
- Servicios sociales

1.13. Formato de los ítems

- Respuesta dicotómica (sí/no, verdadero/falso, etc)
- Tipo Likert
- Se detallan en la siguiente nota

NOTA: La escala de Desarrollo MP-Revisada tiene diferentes tipos de tareas, actividades e ítems. Algunos formatos (y su manera de puntuación) se describen a continuación:

Batería cognitiva (hasta la primera mitad del segundo año). Se utilizan juguetes diversos para provocar las respuestas de los evaluados. En general, los resultados de cada ítem se valoran de forma dicotómica (0 = no realiza la tarea, 1 = sí realiza la tarea). En algunos ítems se controla el tiempo, concediéndose puntos de bonificación si la tarea se realiza en el tiempo establecido.

Batería cognitiva (desde la segunda mitad del segundo año). Los elementos suelen ser estímulos gráficos, aunque también se emplean algunos manipulativos como cubos y rompecabezas. Como en la anterior, la evaluación de los ítems es dicotómica con bonificación por el tiempo en algunos casos.

Motricidad gruesa. El niño tiene que mostrar ciertas habilidades físicas realizando movimientos. Las respuestas también se puntúan de forma dicotómica. Hay además otras dos secciones en las que se recogen observaciones del examinador (calidad del movimiento y movimientos atípicos) que no se puntúan, sirviendo para una valoración cualitativa.

Lenguaje expresivo-Examinador. La escala requiere que el niño exprese en voz alta las respuestas a ciertas tareas y preguntas. Las respuestas se valoran de forma dicotómica.

1.14. Número de ítems

- Los evaluados no reciben todos los ítems de cada escala, sino los apropiados a su edad y nivel. Se indica a continuación los ítems de cada escala y, entre paréntesis, el número de ítems disponibles.

Batería Cognitiva: 49 tareas (226), agrupadas en 7 niveles de dificultad.

Escala de Motricidad gruesa: 23 tareas (77), agrupadas en 5 niveles de dificultad.

Escala de Conducta adaptativa y autocuidado: 43 ítems, agrupados en 5 niveles de dificultad.

Escala Socioemocional: 44 ítems, agrupados en 5 niveles de dificultad.

Escala de Estilo de temperamento: 33 ítems (47), agrupados en 2 niveles de dificultad.

Escala de Lenguaje expresivo. Esta escala consta de dos subescalas. Una de ellas debe ser respondida por los padres y consta de 37 ítems y 8 preguntas sobre antecedentes

comunicativos. La segunda escala está compuesta por 7 tareas de expresión verbal. Presenta 2 niveles de dificultad.

1.15. Soporte

- Administración oral
- Papel y lápiz
- Manipulativo

La escala MP-R requiere de una gran variedad de materiales diferentes en las distintas edades de los niños evaluados, así como administración oral en algunas. También se utilizan cuestionarios de papel y lápiz que deben cumplimentar los padres o cuidadores y el examinador. En la Tabla 2.1 del Manual se recoge una descripción detallada de los materiales necesarios en cada una de las escalas y niveles de edad.

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

- Nivel C

La cualificación es similar a la requerida en otras pruebas de inteligencia y desarrollo de aplicación individual con niños. Las decisiones sobre inicio y finalización de la prueba, así como la necesidad de hacer adaptaciones y modificaciones con determinadas poblaciones, exigen que el examinador sea un psicólogo experto y con experiencia en la evaluación de niños (en especial, cuando ha de evaluar a niños menores de 2 años).

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

- El test es aplicable en edades comprendidas entre 1 mes y 78 meses (6 años y medio); es decir, es aplicable a todos los niños pre-escolares. Los autores del test lo recomiendan especialmente para algunas poblaciones especiales: autistas, niños con deficiencias auditivas, prematuros, y, en general, niños con dificultades de comprensión/expresión en la lengua del test, dado su carácter esencialmente no verbal (salvo, obviamente, las escalas de Lenguaje expresivo).

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

- No existen formas paralelas ni informatizadas.

1.19. Procedimiento de corrección

- Manual.

El examinador, con ayuda de las instrucciones, debe decidir la valoración (0 o 1) de cada uno de los ítems en las escalas de la batería Cognitiva, Motricidad gruesa y Lenguaje expresivo–examinador, así como los puntos de bonificación en los ítems en los que se pueden aplicar. Las instrucciones están claramente especificadas en el Manual. Más detalles de la manera de puntuar cada ítem se han expuesto anteriormente (en el apartado 1.13).

1.20. Procedimiento de obtención de las puntuaciones directas

- Las puntuaciones directas de cada una de las escalas refleja el número de ítems superados correctamente a los que se añaden, en su caso, los puntos de bonificación. Como los sujetos comienzan a responder en diferentes niveles, se calculan las *Puntuaciones totales ajustadas*, que son la suma de los puntos obtenidos por el niño o niña en los ítems aplicados más los puntos posibles antes del punto de inicio. Este procedimiento se sigue en la batería Cognitiva, Índice global, Motricidad gruesa, Conducta adaptativa y autocuidado, y escala Socioemocional.
- En el cuestionario de Estilo de temperamento, cada ítem puntúa 0, 1 o 2, según la frecuencia y se suman las puntuaciones de cada una de las tres escalas que componen el cuestionario. De forma similar se obtienen las puntuaciones directas de las escalas del Comportamiento durante la evaluación.
- El Lenguaje expresivo se evalúa mediante el cuestionario del examinador y el de los padres. Hasta los 12 meses se puede obtener una puntuación combinada: *Puntuaciones infantiles especiales*, uniendo a la de los padres algunos ítems de la batería Cognitiva. A partir de los 13 meses se calcula la *Puntuación en Lenguaje expresivo*, combinando las puntuaciones del Lenguaje expresivo-examinador y Lenguaje expresivo-padres. También se puede obtener un Índice total de lenguaje combinando el Lenguaje receptivo de la batería Cognitiva con el Lenguaje expresivo.
- Pueden corregirse las puntuaciones para el caso de los prematuros teniendo en cuenta el grado de prematuridad.
- No todos los niños reciben los mismos ítems. Al ser una prueba cuyo fin es evaluar el desarrollo desde 0 a 6 años, sus tareas están agrupadas en intervalos de edad, y, dentro de cada uno de ellos, están ordenadas por dificultad creciente. Por esa razón, es necesario establecer para cada niño su “línea base”; es decir, cada niño tiene que realizar un número mínimo de tareas de un nivel con éxito. En caso contrario, se desciende al nivel anterior (“regla de retorno”). Por otra parte, cuando el niño no resuelve correctamente un número determinado de tareas, se da por finalizada la prueba. Para cada escala se establece un procedimiento diferente de estimar “la línea base”, así como reglas específicas de “terminación” de la prueba, o “de retorno” a un nivel inferior.

1.21. Transformación de las puntuaciones

- Normalizada

La transformación se hace a puntuaciones típicas normalizadas salvo en las escalas de Estilo de temperamento y de Comportamiento durante la evaluación.

1.22. Escalas utilizadas

- Centiles
- Puntuaciones típicas en escala de CI (Media =100, DT=15)
- Puntuaciones de desarrollo, obtenidas a partir del modelo de Rasch de la Teoría de la Respuesta al Ítem, y Edades equivalentes.

1.23. Posibilidad de obtener informes automatizados

- No

1.24. Oferta por parte del editor de un servicio para la corrección y/o elaboración informes

- No

1.25. Tiempo estimado para la aplicación del test

- En aplicación individual: Entre 60 y 90 minutos, dependiendo de las escalas que se utilicen.
- En aplicación colectiva: No aplicable

1.26. Documentación aportada por el editor

- Manual
- Vídeo demostración en la web de TEA Ediciones

1.27. Precio de un juego completo de la prueba

- 1017,64 euros.

1.28. Precio y número de ejemplares del paquete de cuadernillos

- 13,52 euros cada paquete de 25 cuadernillos.

1.29. Precio y número de ejemplares del paquete de hojas de respuesta

- No aplicable.

1.30. Precio de la corrección y/o elaboración de informes por parte del editor

- No aplicable.

1.31. Bibliografía básica acerca del test aportada en la documentación

El manual español incluye la bibliografía del manual original y es adecuada. Ofrece una amplia bibliografía relativa a las distintas pruebas de las diferentes escalas y a los procedimientos de medida.

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

Aspectos generales

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Excelente	5
2.2. Calidad de la documentación aportada	Excelente	5
2.3. Fundamentación teórica	Buena	4
2.4. Adaptación del test	Excelente	5
2.5. Calidad de las instrucciones	Excelente	5
2.6. Facilidad para comprender la tarea	Excelente	5
2.7. Facilidad para registrar las respuestas	Buena	4
2.8. Calidad de los ítems (aspectos formales)	Excelente	5
2.9. Datos sobre el análisis de los ítems	Buena	4

- Los Standards for Educational and Psychological Testing señalan que la documentación del test incluirá la naturaleza del test, uso pretendido, proceso seguido en el desarrollo del test, información técnica relacionada a la aplicación, puntuación, interpretación, evidencias de fiabilidad y validez, escalas y normas adecuadas y otros aspectos requeridos para la administración e interpretación del test. El Manual del test es muy exhaustivo en lo que se refiere a todos estos aspectos.
- Dado que el test puede ser aplicado en poblaciones especiales (p.ej., retraso mental, trastornos del espectro autista, deficiencias auditivas, etc.), el Manual proporciona sugerencias sobre las adaptaciones apropiadas en las aplicaciones del test a dichas poblaciones, tal como requieren los citados Standards.
- En la actualidad las ideas dominantes sobre la naturaleza de la inteligencia avalan una estructura jerárquica, con un factor general en el extremo de la jerarquía y múltiples funciones específicas por debajo. Existen variaciones entre los teóricos que defienden el modelo jerárquico, pero la teoría de Carroll de los tres estratos ha sido combinada con la de Cattell-Horn en el modelo Cattell-Horn-Carroll (CHC) que identifica un factor general, amplias habilidades y numerosas habilidades restringidas más específicas. Los autores del test señalan que categorizaron los ítems según el catálogo de factores

cognitivos de Carroll, por lo que puede considerarse dentro de marco CHC. Por otra parte, en las Tablas 9.1 y 9.2 del Manual se presenta la justificación teórica racional de los supuestos procesos subyacentes a cada una de las pruebas.

- Es destacable la inclusión en apéndices de los parámetros de dificultad de los ítems junto con los correspondientes errores típicos. También es destacable el mapa de ítems en el perfil de desarrollo.
- En el Manual se informa de que se realizaron otros análisis de ítems basados en la Teoría clásica de los tests, pero no se presenta información. También se ofrece poca información sobre los estadísticos de ajuste de los ítems al modelo de Rasch.

Validez de contenido

Contenido	Valoración	Puntuación
2.10. Calidad de la representación del contenido o dominio	Buena	4
2.11. Consultas a expertos	Se ha consultado a un número moderado de expertos mediante un procedimiento sistematizado ($10 \leq N \leq 30$)	4

- Se indica en el Manual que se ha consultado a expertos en varias fases del desarrollo de la prueba, pero no se indica específicamente a cuántos.

Validez de constructo

Contenido	Valoración	Puntuación
2.12. Diseños empleados	<ul style="list-style-type: none"> ▪ Correlaciones con otros tests ▪ Diferencias entre grupos ▪ Análisis factorial exploratorio ▪ Análisis factorial confirmatorio ▪ Escalamiento de Rasch y AFE de residuos 	
2.13. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado	4
2.14. Procedimiento de selección de las muestras	Los estudios correlacionales y los estudios con grupos se realizaron con muestras incidentales. El resto de estudios se hizo con la muestra de tipificación, que se obtuvo por un procedimiento de muestreo estratificado cuasi-aleatorio, atendiendo a 3 criterios principales en el diseño del plan de muestreo: edad, sexo y región geográfica.	
2.15. Mediana de las correlaciones del test con otros tests similares	Excelente ($r \geq 0.60$)	5
2.16. Calidad de los test empleados como criterio o marcador	Excelente	5
2.17. Datos sobre el sesgo de los ítems	No se aporta	-

Comentarios sobre la validez de constructo

- Es destacable la multiplicidad de procedimientos seguidos para establecer las evidencias de validez. No obstante, en la parte del Manual en la que se presentan los análisis factoriales exploratorios y confirmatorios, sería conveniente diferenciar mejor los de la versión original y los de la adaptación española.
- En la versión original se utilizaron en el estudio correlacional las Escalas Bayley de desarrollo infantil, la escala Leiter-R y la Escala Stanford-Binet 5ª ed. (SB5). En la adaptación española se utilizaron como referentes el Inventario de desarrollo de Batelle y las escalas de McCarthy de Aptitudes y Psicomotricidad. Las correlaciones con las subescalas del test Batelle son todas superiores a 0.70, tanto en el grupo de niños con desarrollo típico, como en el grupo clínico. Las correlaciones con las subescalas del

McCarthy superan en general el valor de 0.60, excepto en las escalas cumplimentadas por los padres.

- El Manual indica que en la muestra americana se ha estudiado el sesgo en relación a sexo y grupos étnicos. Los procedimientos utilizados no son los habituales en los estudios de funcionamiento diferencial y se dan pocos detalles de los resultados alcanzados. En relación a la muestra española, no consta que se hayan realizado estudios de funcionamiento diferencial.

Validez predictiva

- En las relaciones con criterios utilizan una aproximación concurrente, basada en el análisis de las diferencias entre grupos clínicos. En el estudio original en USA utilizaron muestras pequeñas (entre 14 y 43 casos, según el grupo) de cada uno de los siguientes grupos: retraso mental, prematuros, retraso del habla y lenguaje, sordera o dificultades auditivas severas, retraso o desviación motora y trastorno del espectro autista. En la adaptación española participaron 48 casos con distintos trastornos y los resultados muestran las mismas tendencias que los obtenidos con el test original.

Contenido	Contenido	Puntuación
2.18. Diseño de selección del criterio	<ul style="list-style-type: none"> ▪ Concurrente ▪ Retrospectivo 	
2.19. Tamaño de las muestras	Un estudio con una muestra moderada ($100 \leq N < 200$)	2
2.20. Procedimiento de selección de las muestras		
2.21. Mediana de las correlaciones del test con los criterios	No se aporta	-

Comentarios sobre validez predictiva

- En los estudios realizados con el test original aplicaron varias escalas del test a distintos grupos clínicos (retraso mental, prematuros, trastornos de habla y lenguaje) y a grupos sin trastorno. Obtuvieron los siguientes estadísticos: aciertos en la clasificación (entre 78,3 y 99,6%), especificidad (entre 81,5 y 95,8%), sensibilidad (39,1 a 95,8%), falsos positivos (de 0,2 a 17,1%) y falsos negativos (entre 0,2 y 4,6%).

Comentarios generales sobre validez

Las evidencias de validez han sido abordadas desde casi todos los ámbitos recomendados en la literatura psicométrica, lo que puede considerarse un punto fuerte de la escala. Las correlaciones con las puntuaciones de otros tests similares son elevadas y los estadísticos basados en la capacidad de discriminar entre grupos clínicos y no clínicos alcanzan valores muy aceptables. El Manual informa además, como parte del estudio de validez de contenido, de las capacidades que requiere cada tarea.

Se analizaron también las evidencias de validez como estructura interna. En la adaptación española se realizaron análisis factoriales exploratorios y confirmatorios de las pruebas cognitivas. También se abordó el estudio de la unidimensionalidad mediante el análisis de los residuos del modelo de Rasch. En la adaptación española también se analizó la estructura del cuestionario de Temperamento y la de Comportamiento durante la evaluación.

En la adaptación española también es interesante el análisis del crecimiento de las puntuaciones que se produce en el Índice global con incrementos en la edad.

Creemos que podría mejorarse en el manual la presentación y explicación de los resultados de los diversos análisis factoriales, especialmente los realizados a partir de las puntuaciones cognitivas.

Fiabilidad

Contenido	Valoración	Puntuación
2.22. Datos aportados sobre fiabilidad	<ul style="list-style-type: none"> ▪ Coeficientes de fiabilidad para diferentes grupos de sujetos ▪ Error típico de medida para diferentes grupos de sujetos 	

Equivalencia formas paralelas

Contenido	Valoración	Puntuación
2.23. Tamaño de las muestras	No se aporta	-
2.24. Coeficientes de equivalencia	No se aporta	-

En este tipo de pruebas, de elaboración de ítems muy costosa, no es frecuente la creación de formas paralelas.

Consistencia interna

Contenido	Valoración	Puntuación
2.25. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado	4
2.26. Mediana de los coeficientes de consistencia	Excelente ($r \geq 0.85$)	5

- Los estudios de consistencia interna se llevaron a cabo para cuatro grupos diferentes de edad, con muestras de tamaños entre $n = 124$ y $n = 374$, y para cada una de las escalas, así como para el Índice global.

Estabilidad (test-retest)

Contenido	Valoración	Puntuación
2.27. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado	4
2.28. Coeficientes de estabilidad	Excelente ($r \geq 0.80$)	5

Comentarios generales sobre fiabilidad

- En el caso de las escalas de desarrollo y, en general, de los test para la evaluación de pre-escolares, se recomienda considerar dos tipos de coeficientes de fiabilidad: consistencia interna y estabilidad temporal mediante test-retest. Los dos tipos de coeficientes son calculados en el test MP-R para el Índice global, las diferentes escalas

y en cuatro grupos de edad. En el caso de la consistencia interna, podemos considerar que la fiabilidad es excelente.

- Se aportan los errores típicos de medida de cada una de las escalas y para cada grupo de edad, como recomiendan los Standards, con objeto de poder construir intervalos de confianza para las puntuaciones. Como cabe esperar, el error de medida es mucho mayor en los dos extremos del intervalo de aplicación de la prueba.
- En el caso de la fiabilidad como estabilidad temporal de las puntuaciones, los valores de los coeficientes de correlación de Pearson convencionales y corregidos por la restricción del rango son también elevados.

Baremación

Contenido	Valoración	Puntuación
2.29. Calidad de las normas	Varios baremos dirigidos a diversos estratos poblacionales	4
2.30. Tamaño de las muestras	Grande ($600 \leq N < 1000$)	4
2.31. Procedimiento de selección de las muestras	Se ha aplicado un procedimiento de muestreo estratificado cuasi-aleatorio, atendiendo a 3 variables demográficas (sexo, edad y región) para que la proporción de niños en la muestra de tipificación sea básicamente la que existe en el censo.	

Los baremos contemplan varios estratos de edad con intervalos de solo uno, dos y tres meses.

- La MP-R proporciona puntuaciones típicas, percentiles, equivalentes de edad y puntuaciones de desarrollo (criteriales). Las puntuaciones pueden representarse gráficamente en un Perfil de Desarrollo desglosado en: Índice Global, Cognición, Motricidad fina, Lenguaje receptivo y Motricidad gruesa.
- El Perfil de puntuaciones de desarrollo permite al examinador representar gráficamente las puntuaciones de los cinco dominios relacionándolos con los equivalentes de edad. Pueden identificarse áreas deficitarias por los ítems del test

fallados, lo que permite planificar intervenciones educativas para mejorar esos déficits.

- La muestra de tipificación pretende lograr la representatividad por medio de la adecuación a variables del censo, con cuotas por sexo, edad y región. Está formada por $N = 968$ niños. Aproximadamente se han recogido 150 casos para cada año de edad. Para la construcción de los baremos la muestra se ha segmentado en numerosos grupos y se ha aplicado la técnica de *continuous norming*, que tiene en cuenta la muestra completa en la obtención de los baremos de cada grupo. Al aplicar esta técnica se consigue reducir el impacto del menor número de evaluados de los grupos en relación a la muestra completa.

Comentarios generales sobre baremación

a) Sobre el grupo normativo.

En el caso de los baremos o normas de interpretación para las escalas de desarrollo, Alfonso y Flanagan (2008) señalan que hay que valorar los siguientes aspectos del grupo normativo: el tamaño del grupo normativo en cada edad y/o grado, el tiempo transcurrido desde que se hicieron las evaluaciones, el rango de edad de las normas, y el emparejamiento entre las características demográficas de la muestra y de la población. De los cuatro aspectos anteriores consideramos muy adecuados los tres últimos. En cuanto al primero, pensamos que vendría bien aumentar el tamaño de las muestras utilizadas en las edades más tempranas, de 0 a 11 meses, pues es el grupo de edad con un menor número de evaluados cuando la técnica *continuous norming*, aplicada a la construcción de los baremos, recomienda que haya más evaluados en los grupos más extremos.

b) Sobre los baremos.

El Manual ofrece baremos para grupos de edad de rango muy reducido (1, 2 y 3 meses) y moderado tamaño. Para solventar esta dificultad, se ha recurrido muy acertadamente a la técnica *continuous norming*, que tiene en cuenta la muestra total en la confección de cada baremo específico.

El manual del test presenta diversas posibilidades de interpretación:

- 1) Referida a las normas: puntuaciones típicas derivadas (en escala de media = 100 y desviación típica = 15) y percentiles.
- 2) Referidas a criterio, utilizando *puntuaciones de desarrollo* procedentes de un escalamiento con el modelo de Rasch dicotómico. Existen puntuaciones de desarrollo para todas las escalas y un examen de las mismas hace pensar que el rango es adecuado, sin efectos suelo ni techo, que son dos fallos que se deben evitar en las escalas evolutivas. Este tipo

de escalas contiene numerosos puntos por lo que pueden considerarse bastante sensibles al cambio y que las hacen recomendables para la evaluación de intervenciones.

3) Edades equivalentes obtenidas a partir de las escalas anteriores.

c) Otros aspectos de interés de la interpretación de las puntuaciones.

Se sugieren otras posibilidades de interpretación de las puntuaciones que pueden proporcionar información relevante para el diagnóstico, como las puntuaciones ajustadas en el caso de los niños prematuros, cálculo de la discrepancia entre diferentes escalas en términos de puntuaciones típicas y de puntuaciones de desarrollo, cálculo de los porcentajes de retraso o discrepancia entre el resultado esperado según la edad y el observado o magnitud del retraso, la relación de ítems dominados y fallados (lo que ayuda en la planificación de la intervención de niños con retrasos), etc.

Conviene destacar además el buen diseño de la *Hoja resumen de resultados y perfil de desarrollo* que facilita enormemente la tarea de la puntuación e interpretación.

Referencias

Alfonso, V. C., & Flanagan, D. P. (2008). *Assessment of preschool children: A framework for evaluating the adequacy of the technical characteristics of norm-referenced instruments*. New York: Wiley.