

Evaluación del

Perfil de Desarrollo - 3

∞ DP - 3 ∞

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Perfil de Desarrollo – 3 (DP-3)
Autor	Gerald D. Alpern
Autor de la adaptación española	Fernando Sánchez Sánchez (Dpto. de I+D+i de TEA Ediciones)
Variable(s)	Nivel de desarrollo (comunicación, conducta adaptativa, cognición, socioemocional, motricidad)
Áreas de aplicación	Psicología educativa, psicología clínica, neuropsicología, investigación.
Soporte	Administración oral, papel y lápiz, informatizada. Corrección por internet

Valoración general

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

Comentarios generales

El Perfil de Desarrollo – 3 (DP-3) es la adaptación española de la última versión del instrumento original Developmental Profile – II (Alpern, Boll y Shearer, 1986). Los constructos que evalúa y su estructura teórica están basados en el modelo multidimensional del desarrollo infantil de Alpern y Boll (1972), que contempla cinco grandes áreas de desarrollo: motricidad, lenguaje y comunicación, desarrollo socioemocional, desarrollo cognitivo y conducta adaptativa. Se trata de una prueba de cribado o 'screening' que permite evaluar de forma rápida y detectar posibles retrasos en cinco áreas del desarrollo infantil (cognición, motricidad, socioemocional, comunicación y conducta adaptativa), ofreciendo un índice global de desarrollo de la persona evaluada. Si bien es necesario ampliar los estudios sobre su sensibilidad y especificidad, parece discriminar con suficiente precisión un proceso de desarrollo 'normal' de uno en el que pueda darse algún tipo de alteración. Puede utilizarse en diversos ámbitos con diversos fines, si bien se hace hincapié en que no debe utilizarse de forma aislada para hacer diagnósticos ni establecer tratamientos, teniéndose que interpretar los resultados atendiendo al entorno familiar, cultural y social de la persona evaluada y a las oportunidades que esta ha tenido para alcanzar un determinado nivel de desarrollo.

Consta de 193 ítems que son aplicados atendiendo al momento de desarrollo en que se encuentra la persona evaluada y siguiendo unas reglas de retorno y parada que permiten determinar la línea de base y el fin de la aplicación al haberse alcanzado el límite superior del nivel de desarrollo actual de la persona evaluada. El instrumento cuenta con dos versiones, una en formato de entrevista y otra que es cumplimentada directamente por los progenitores o cuidador. Es una prueba de rápida cumplimentación (entre 20 y 40 minutos, si bien el cuestionario puede llevar algo más de tiempo al requerir que los padres contesten a todas las preguntas del test).

Su corrección e interpretación resulta bastante sencilla al estar informatizada y ofrecer puntuaciones con las que los profesionales están familiarizados (puntuaciones directas, percentiles y puntuaciones CI). El sistema de corrección permite obtener dos informes narrativos diferenciados para el profesional y la familia que resultan útiles al contener estrategias para trabajar los puntos débiles identificados en el niño o la niña y estar adaptados específicamente para que sean comprensibles por los receptores. Todas estas características hacen del DP-3 un instrumento interesante para ser utilizado en el ámbito educativo ante la sospecha o aparición de señales que alarmaran sobre un posible retraso o trastorno del desarrollo.

Los materiales y documentación proporcionados son de una calidad excelente, con un formato cuidado y práctico, lenguaje inclusivo e información detallada acerca de la mayor parte de los aspectos esenciales de una prueba psicológica. La facilidad para responder en los cuadernillos también se valora muy positivamente. Asimismo, se proporciona información detallada y adecuada acerca del desarrollo de la versión original del DP-3 y sobre el proceso de adaptación española, que ha seguido las recomendaciones de la International Test Commission (2010). La bibliografía se encuentra muy actualizada en lo que respecta a la metodología y los estudios realizados con la adaptación española, observándose una menor actualización en la referida a las áreas de desarrollo, si bien esta circunstancia parece explicarse por tratar de respetar la bibliografía utilizada en el instrumento original estadounidense.

Los baremos, análisis y resultados presentados se llevan a cabo sobre una muestra depurada muy amplia (N=2.227) y heterogénea, siendo el 79% de los casos extraídos aleatoriamente de la población general y el 21% restante de manera incidental en una muestra clínica. Esta es precisamente otra de las fortalezas del instrumento: la utilización de dos muestras de considerable tamaño, una clínica y otra de población general obtenida esta mediante un cuidadoso muestreo para garantizar su representatividad. Del mismo modo, se valora muy positivamente la utilización de la metodología de baremación continua ('continuous norming'; Lenhard et al., 2016) para la construcción de los baremos y la combinación de técnicas de la Teoría Clásica de los Tests (TCT), Teoría de Respuesta a los Ítems (TRI) y Modelos de Ecuaciones Estructurales para el desarrollo y validación de la prueba.

Las evidencias de fiabilidad se basan en el análisis de la consistencia interna, la estabilidad temporal, los errores típicos de medida y los intervalos de confianza. Se considera otro punto fuerte que la fiabilidad se ha estudiado en muestras amplias y variadas, tanto clínicas como de población general, proporcionándose información detallada para diversos grupos.

Las evidencias de validez recopiladas sobre el DP-3 incluyen aquellas relacionadas con el contenido del test, con su estructura interna, con las relaciones con otras variables externas y con las consecuencias de la evaluación. Se consideran puntos fuertes de estas evidencias la cuidada lista final de ítems (modificando y reemplazando algunos de los originales y añadiendo otros nuevos), los indicadores de fiabilidad obtenidos, las evidencias de validez de contenido proporcionadas y los 29 baremos atendiendo a la edad de la persona evaluada. También se valora muy positivamente el que se hayan realizado diversos estudios de validez con población española, además de los realizados con versiones previas estadounidenses, incluyendo grupos clínicos amplios y variados. Del mismo modo, se consideran puntos fuertes de la prueba los estudios correlacionales con otros tests relevantes en el área.

En términos generales, las propiedades psicométricas del test se consideran buenas (excelentes en varias de las áreas evaluadas), si bien en futuras ediciones sería deseable tener en cuenta los siguientes aspectos:

- Aunque se aportan suficientes y adecuadas evidencias sobre la estructura interna de la escala (análisis factorial exploratorio, confirmatorio, correlaciones interescales y correlaciones ítem-escala), conviene aclarar que el análisis factorial confirmatorio (AFC) realizado proporciona evidencias del ajuste de un modelo unidimensional a los datos, cuando la prueba proporciona puntuaciones en cinco escalas y un índice general. En este sentido, sería deseable comparar el ajuste de esta solución con otras más acordes con el modelo teórico en el que se sustenta.
- Los Errores Típicos de Medida oscilan de 2.13 a 8.69 puntos en la versión de entrevista y de 1.21 a 7.44 en la versión cuestionario. En el Anexo C se indican los errores máximos de medida para el cálculo de los intervalos de confianza, resultando más elevados en población clínica de corta edad y en población general a partir de los 10 años, algo que se advierte en el Manual e implica que las puntuaciones de estos grupos deban interpretarse con cautela.
- Dado que se ha utilizado TRI en la calibración de los ítems, sería recomendable incluir información sobre la Función de Información de cada uno de los ítems del test pues, teniendo en cuenta que se trata de una prueba de 'screening' de posibles retrasos en el desarrollo, interesa medir con más precisión en la zona de habilidad comprendida en torno a la media y dos desviaciones típicas por debajo de esta.
- En la muestra clínica existe cierta sobrerrepresentación de trastornos del neurodesarrollo (TDAH, TEA y DI) frente a otro tipo de problemas de aprendizaje o alteraciones, lo que lleva a aconsejar que el DP-3 sea especialmente recomendado en el cribado de este tipo de condiciones.
- En cuanto al cuestionario, debiera contemplarse alguna vía para que los padres de un/a niño/a, por ejemplo, de 8 años, no tengan que responder "sí" a las preguntas apropiadas para aquellos más pequeños. Algo similar ocurriría cuando se aplica a un niño de 1 año, pues se obliga al informante a responder "no" a muchas preguntas que son apropiadas para personas de más edad. Cabría explorar si puede generarse versiones adaptativas del cuestionario y la entrevista para conseguir que los informantes respondan solo a las preguntas realmente informativas del nivel de desarrollo. Los editores parecen ser conscientes de la dificultad e insisten en el cuadernillo y en manual en la necesidad de que se responda a todo, aunque pueda parecer que no tiene sentido tal exigencia.
- En próximas ediciones pueden corregirse algunas erratas de las páginas(lineas) siguientes: 35(-6), 59(19), 140 (-7 a -5), 145 (tabla 6.24) y 230 (escala Cog).

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

Perfil de Desarrollo - 3 (DP-3).

1.2. Nombre del test en su versión original:

Developmental Profile – 3 (DP-3).

1.3. Autor/es del test original:

Gerald D. Alpern.

1.4. Autor/es de la adaptación española:

Fernando Sánchez Sánchez (Dpto. de I+D+i de TEA Ediciones).

1.5. Editor del test en su versión original:

Western Psychological Services (WPS).

1.6. Editor de la adaptación española:

TEA Ediciones.

1.7. Fecha de publicación del test original:

2007 (3ª edición).

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

2018.

1.9. Fecha de la última revisión del test:

2018.

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

Escalas de desarrollo.

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

El perfil de desarrollo DP-3 es un instrumento de evaluación dirigido a valorar diversas áreas relevantes en el proceso de desarrollo del menor (desde los 0 hasta los 12 años y 11 meses) con el fin de identificar posibles retrasos y ofrecer una intervención lo más temprana posible. La información sobre las habilidades y aptitudes del niño es proporcionada por sus progenitores o las personas a cargo del cuidado, bien a partir de la entrevista con un profesional, bien mediante autoinforme, contando por tanto con dos versiones diferentes, pero de idéntico contenido, que permiten adaptar la evaluación a la disponibilidad de las madres o los padres que responden a la prueba. Además, el cuestionario se puede aplicar on-line.

La estimación del nivel del desarrollo infantil se realiza a través de la evaluación de cinco áreas:

- **Motricidad:** mide el desarrollo de la motricidad y contribuye a determinar la capacidad del niño o de la niña para desarrollar tareas que requieren coordinación muscular, fuerza, flexibilidad y otras habilidades motoras. Incluye elementos para evaluar motricidad fina y motricidad gruesa.
- **Conducta adaptativa:** esta escala es similar a la escala Autocuidado que aparecía en anteriores versiones del DP. Evalúa la capacidad y la madurez para enfrentarse a las exigencias del entorno. Se refiere a la competencia en tareas como alimentarse, vestirse, actuar de forma independiente y utilizar la tecnología actual.
- **Socioemocional:** esta escala es similar a la escala Social de las versiones anteriores del DP. Evalúa las capacidades relativas a las relaciones interpersonales, la comprensión de aspectos sociales y emocionales, y la forma de actuar en determinadas situaciones sociales. Más concretamente, se centra en evaluar la forma en que la persona se relaciona con sus amigos, familiares y otros adultos.
- **Cognición:** esta escala es similar a la de Rendimiento académico de las versiones anteriores del DP. Evalúa de forma indirecta la capacidad cognitiva a través de las habilidades de desarrollo necesarias para conseguir un rendimiento académico e intelectual adecuado (no mide por tanto inteligencia o rendimiento de forma directa). En las primeras etapas, evalúa las habilidades previas necesarias para conseguir un buen rendimiento escolar mientras en los niveles superiores se evalúan capacidades escolares reales.
- **Comunicación:** evalúa la capacidad de comunicación expresiva y receptiva tanto de forma verbal como no verbal. Incluye aspectos como la comprensión, la expresión oral, el lenguaje gestual y la capacidad de comunicarse eficazmente mediante el uso de determinados instrumentos como el teléfono o el ordenador.

Consta de un total de 193 ítems que en la versión entrevista no se aplicarán en su totalidad, sino en función de la edad del niño o la niña. La aplicación es individual y dura aproximadamente 35 minutos (tanto la entrevista como el cuestionario).

Las puntuaciones tipificadas permiten conocer si el desarrollo de la persona evaluada se encuentra por encima o por debajo del punto de corte, ofreciendo puntuaciones de fácil interpretación para cada área (directas, centiles y puntuaciones transformadas a escala CI), así como un índice general de desarrollo, que se obtiene combinando las puntuaciones típicas de las cinco escalas principales.

Se proporcionan baremos en función de la edad, en tramos de dos meses para los menores de 2 años, en tramos de cuatro meses para los de 3 años, en tramos de seis meses para los de 4-6 años y en tramos de un año para los de 7-12 años. De este modo, el instrumento permite identificar los puntos fuertes y débiles del/a niño/a en las áreas mencionadas, por lo que la información ofrecida por el DP-3 puede ser muy relevante tanto en el ámbito clínico a modo de 'screening' de distintos tipos de trastornos del neurodesarrollo o problemas de aprendizaje, como educativo (p. ej., desarrollo de adaptaciones curriculares que respondan a las necesidades y aptitudes de la persona).

1.12. Áreas de aplicación:

Psicología clínica, psicología educativa, neuropsicología, investigación.

1.13. Formato de los ítems:

Respuesta dicotómica (sí/no).

1.14. Número de ítems:

193 en total, organizados en cinco escalas formadas por entre 35 y 42 ítems, ordenados de menor a mayor dificultad.

Motricidad= 40 ítems.

Conducta adaptativa= 40 ítems.

Socioemocional= 36 ítems.

Cognición= 42 ítems.

Comunicación= 35 ítems.

En la versión entrevista se aplica solo un conjunto de ítems en función de la edad de la persona evaluada y de su nivel de desarrollo en cada área (es necesario establecer la línea base en cada escala). En la versión cuestionario se han de contestar todos los ítems.

1.15. Soporte:

Administración oral, papel y lápiz. Informatizada. Corrección por internet.

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

Nivel B.

El DP-3 puede utilizarse indistintamente tanto en su formato entrevista como cuestionario, aunque cada uno de ellos requiere nivel de experiencia y entrenamiento diferentes. En el caso del cuestionario, puede ser aplicado por cualquier persona que disponga de suficiente entrenamiento en la utilización de cuestionarios y tests psicológicos, familiarizada con el contenido y procedimiento de aplicación del DP-3. La entrevista, por su parte, es fácil de aplicar y corregir por una persona con experiencia en evaluación psicológica, aunque también puede ser aplicada por otros profesionales con el debido entrenamiento.

La interpretación y acceso a los resultados, en cambio, se reserva a profesionales titulados y cualificados con conocimiento sobre teoría de los tests, métodos estadísticos y evaluación de aspectos del desarrollo infantil (p. ej., psicólogos clínicos, neuropsicólogos o psicólogos educativos).

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

Población infantil con edades comprendidas entre los 0 años y 0 meses y los 12 años y 11 meses que presenten tanto un desarrollo normal como sospecha de cierto retraso en alguna de las áreas evaluadas por el test o general.

Puede aplicarse a grupos clínicos. De hecho, en la adaptación española, el test se aplicó a niños diagnosticados de, según el DSM, discapacidad intelectual, autismo y otros trastornos (del aprendizaje, lenguaje y comunicación, motores, déficit de atención e hiperactividad, y emocionales).

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

Existe una versión de entrevista para ser aplicada por un profesional a padres, madres o cuidadores y otra versión cuestionario que puede ser cumplimentada directamente por estos últimos (i.e., autoinforme). Existe la posibilidad por tanto de aplicar el instrumento de manera oral (entrevista), escrita (autoinforme) e incluso on-line a través de la plataforma de TEA. Para hacer uso de la plataforma online es necesario el registro, así como la disposición de las claves o pines de corrección.

1.19. Procedimiento de corrección:

La corrección es automatizada, mediante una plataforma de corrección on-line.

1.20. Puntuaciones:

La prueba se puede aplicar como entrevista y como cuestionario. Los ítems, dicotómicos (sí/no), todos ellos directos, son los mismos en ambas aplicaciones.

En el cuestionario, el informante (la madre, padre o tutor/a) ha de responder a los 193 ítems. En la entrevista, se plantea al informante las preguntas apropiadas, según la edad de la persona evaluada. Ha de prestar especial atención al punto de inicio, regla de retorno (a aplicar si el punto de inicio no fuese el correcto) y la regla de parada. Al hacerlo así, se espera obtener una puntuación directa similar a la obtenida en el cuestionario, aplicando solo los ítems realmente informativos (evitando los que van a producir muy probablemente la respuesta "sí" o la respuesta "no").

Las puntuaciones directas constituyen la suma de las respuestas afirmativas o 'síes' a cada uno de los ítems. Estas son después transformadas, a través de la plataforma de corrección on-line, a una distribución CI (puntuaciones por tanto comparables entre subescalas y con respecto al índice general de desarrollo calculado) contando también con puntuaciones centílicas (se incluye, en el Anexo D, la correspondencia entre varias escalas típicas transformadas y los percentiles).

En el Anexo C se indican los errores máximos de medida para el cálculo de los intervalos de confianza, resultando llamativamente elevados, especialmente amplios en población clínica de corta edad y en población general a partir de los 10 años, algo que se advierte en el Manual, aunque sólo para las puntuaciones extremas. También se incluye un anexo para determinar si las diferencias entre puntuaciones en las distintas áreas evaluadas por el DP-3 son significativas (Anexo E).

Con respecto al control de patrones de respuesta inconsistentes, se incluye una escala de control sólo en el formato de autoinforme (no en la entrevista). El nivel de inconsistencia de las respuestas es valorado a partir de esta escala atendiendo a tres indicadores: Alerta (el cuestionario no es interpretable), Precaución (se recomienda cierta cautela a la hora de interpretar los resultados) o Normal.

La corrección no se puede realizar manualmente. Las respuestas dadas por los respondientes deben introducirse en la plataforma de corrección por internet (www.teacorrige.com) y el sistema envía un perfil individual con los resultados obtenidos.

1.21. Escalas utilizadas:

Puntuaciones basadas en percentiles (centiles) y puntuaciones CI de desviación ($M=100$; $DT=15$).

Además, en el Anexo D, se incluye una tabla con la correspondencia entre diferentes escalas transformadas para aquellos usuarios que deseen expresar los resultados obtenidos en otras escalas de medida: eneatis (M=5, DT=2), decatipos (M=5; DT=2), escala S (M=50; DT=20); escala T (M=50; DT=15).

1.22. Posibilidad de obtener informes automatizados:

Sí. El programa informatizado devuelve un informe dirigido al profesional y otro destinado al familiar.

En el primero se incluyen las puntuaciones (directas, centiles y CI) obtenidas por la persona evaluada en cada una de las áreas (también brevemente descritas en el informe) junto con un perfil de desarrollo. Aparece, asimismo, la puntuación del Índice General de Desarrollo (IGD). En la versión cuestionario aparece también la puntuación de la Escala de Inconsistencia (que informa de la calidad de las respuestas, si ha habido respuestas dadas con poca atención, etc.), clasificada en tres niveles: normal, precaución y alerta. Además, se proporciona una breve interpretación de las puntuaciones obtenidas por la persona evaluada en el IGD y en cada una de las áreas y se indica la posible existencia de diferencias estadísticamente significativas entre las mismas, con ánimo de identificar puntos fuertes y débiles de la persona evaluada. A continuación, se indican los ítems que están por debajo de su nivel de desarrollo pero que no han sido superados, con el objetivo de facilitar una intervención posterior, para lo cual se proponen algunas actividades.

El informe dirigido a la familia omite las puntuaciones incluyendo sólo el perfil gráfico del menor y describe, con un lenguaje menos técnico y sin incidir en puntuaciones, el rendimiento de la persona evaluada en cada área y en el IGD, señalando actividades a realizar por los padres para mejorar las habilidades en las que se observa un posible retraso.

El perfil gráfico permite realizar un análisis rápido e intuitivo de las puntuaciones. El gráfico se sombrea con diferentes colores para indicar diferentes niveles: naranja para puntuaciones inferiores a la media (CI=70-84) y naranja oscuro para puntuaciones muy inferiores a la media (CI < 70), con el punto de corte para la remisión representado con una línea punteada situada a -1,5 DT (CI=77).

Opcionalmente se puede solicitar un informe narrativo, que proporciona un detallado documento (más de 10 páginas) con: (a) una interpretación de cada una de las 6 escalas (las 5 y la general), (b) observaciones sobre las diferencias encontradas entre las distintas áreas evaluadas, (c) detalles sobre las respuestas dadas a algunos ítems específicos, y (d) algunas recomendaciones o propuestas de intervención.

1.23. Tiempo estimado para la aplicación del test:

35 minutos aproximadamente, en aplicación individual:

20-40 minutos, la entrevista

30-40 minutos, el cuestionario.

1.24. Documentación aportada por el editor:

Manual, cuadernillo del cuestionario, cuadernillo de la entrevista y clave de acceso (PIN) para la corrección por internet.

1.25. Precio de un juego completo de la prueba:

124,80 euros (Manual, 10 ejemplares de cada tipo y 1 PIN de corrección 20 usos).

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

52,94 euros (25 ejemplares del cuestionario o la entrevista, PIN de corrección 25 usos)

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:

52,94 euros (PIN de 25 usos).

115,53 euros (10 e-informes).

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	Excelente	5
2.3. Fundamentación teórica	Buena-Excelente	4,5
2.4. Adaptación del test	Excelente	5
2.5. Desarrollo de los ítems del test	Buena-Excelente	4,5
2.6. Calidad de las instrucciones para el participante	Buena-Excelente	4,5
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	Adecuada-Buena	3,5
2.10. Datos sobre el análisis de los ítems	Excelente	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	Buena	4
2.11.1.2. Consultas a expertos	Se ha consultado a un amplio número de expertos	5

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 otros tests, diferencias entre grupos	
2.11.2.1.2. Tamaño de las muestras	Varios estudios con muestras clínicas 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	Buena	4
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	Buena	4
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	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	Comparación de puntuaciones en diferentes grupos clínicos con la versión original y la adaptada	
2.11.2.2.2. Calidad de los criterios empleados	Buena	4
2.11.2.2.3. Relación temporal entre test y criterio	Concurrente	
2.11.2.2.4. Tamaño de las muestras	Varios estudios con muestras clínicas de tamaño moderado	4
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental y aleatorio, aunque las muestras no son representativas de la población objetivo	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	No se aporta	---

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Excelente	5
2.11.3.2. Funcionamiento diferencial de los ítems	Buena	4

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:

Son diversos los indicadores de validez que se ofrecen en el manual, tanto para la versión entrevista como para la versión cuestionario de la adaptación española (omitimos aquí datos referidos a la versión original o versiones previas del instrumento, también referidos en el manual).

Con respecto a la validez de contenido, si bien se echa en falta algo más de información sobre la selección y análisis del contenido del banco de ítems de la versión original, la adaptación española incluye un procedimiento sistematizado de consulta a 52 expertos que pone de manifiesto que los ítems del DP-3 representan el constructo que se pretende medir.

Se aportan asimismo suficientes evidencias sobre la estructura interna de la escala (análisis factorial exploratorio, confirmatorio, correlaciones interescales y correlaciones ítem-escala). Sobre este aspecto, aunque las evidencias aportadas se valoran de forma muy positiva, llama la atención en primer lugar que se lleve a cabo análisis factorial exploratorio (AFE) y confirmatorio (AFC) sobre la misma muestra. Dado que la prueba se basa en un modelo teórico ampliamente aceptado y se trata de la tercera edición de un instrumento ampliamente testado, parece más conveniente realizar solo AFC. Sobre este, hubiera sido interesante disponer de una explicación más detallada acerca de la decisión de utilizar las cinco escalas (no los ítems) como indicadores y que se hubiese explorado el ajuste a los datos de otras soluciones factoriales alternativas. Del mismo modo, atendiendo a la estructura de la prueba, parece recomendable el uso de un modelo bifactor.

Con respecto a la validez convergente, los datos sobre dos muestras de pequeño tamaño en las que se aplican el MP-R y el Inventario de Desarrollo Battelle junto con el DP-3 tanto en versión entrevista como en versión cuestionario (N=106 y N=42, respectivamente) indican correlaciones cuyo valor promedio puede considerarse bueno en el caso del MP-R (Mdn de los coeficientes de .63 y .60 en entrevista y cuestionario, respectivamente) y adecuado en el caso del Inventario de Desarrollo Battelle (Mdn de los coeficientes de .58 y .57 en entrevista y cuestionario, respectivamente). Estos datos son, no obstante, interpretados en el Manual como indicativos de una validez convergente excelente para el MP-R y buena en el caso del Inventario Battelle. No se han llevado a cabo estudios en los que se examine la correlación del DP-3 con otros instrumentos con los que no debiera estar relacionado.

Para analizar la validez referida al criterio, se realizan comparaciones entre una amplia muestra clínica (N=469) con distintos perfiles de alteraciones en el desarrollo y una muestra de comparación de población general pareada en edad y sexo (N=469). También se analizan las diferencias entre las diferentes submuestras, siendo algunas de reducido tamaño (oscilan desde los 19 -nacimiento prematuro- a los 100 casos -trastornos del espectro del autismo-). Se aportan evidencias de la capacidad de la prueba para detectar diferencias estadísticamente significativas entre las puntuaciones CI promedio de la muestra general y la/s muestra/s clínica/s. Cabe destacar que se ofrecen resultados con el test original y con la versión adaptada, obteniéndose un patrón de resultados similar con ambas muestras y con las dos versiones del test (cuestionario/entrevista). Por otro lado, hubiera sido deseable, dado el objetivo del test, analizar la sensibilidad y especificidad del instrumento de evaluación, al contar con casos con diagnósticos concretos que hubieran permitido analizar la capacidad de la prueba para identificar correctamente a aquellos que, con sospecha de un retraso en el desarrollo, recibirán un diagnóstico posterior.

Además, se ha realizado un estudio del funcionamiento diferencial, en relación con el sexo, y se ha comprobado el funcionamiento de la escala de inconsistencia (que pretende descubrir patrones de respuesta azarosos e inconsistentes cuando se responde al cuestionario).

Finalmente, un aspecto destacable es que se hayan estudiado las posibles diferencias entre las puntuaciones obtenidas con la versión de entrevista y con la versión de cuestionario, pues la comparación de los resultados obtenidos por una persona evaluada con ambas versiones supone una fuente adicional de evidencias de la validez del instrumento. En la adaptación española, se evaluó con las dos versiones prácticamente a la totalidad de la muestra de tipificación y de la muestra clínica (más de 2.200 participantes), sin observarse diferencias significativas sino una alta correlación entre las medias obtenidas con las distintas versiones. Estos resultados, por tanto, pueden interpretarse como una elevada convergencia entre los resultados obtenidos mediante una u otra versión. No obstante, ha de notarse que el hecho de que la media no difiera y la correlación sea muy alta no implica necesariamente que dos puntuaciones de una persona, obtenidas con la entrevista y el cuestionario, sean idénticas. A este respecto, se consideran muy interesantes los tests de verificación surgidos en el ámbito de la Psicología de la Organizaciones que permiten contrastar específicamente si dos versiones de un test asignan a una persona la misma puntuación.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Coeficientes de fiabilidad y error típico de medida para diferentes grupos de personas	

2.12.2. Equivalencia formas paralelas:

Contenido	Valoración	Puntuación
2.12.2.1. Tamaño de las muestras	---	---
2.12.2.2. Puesta a prueba de los supuestos de paralelismo	---	---
2.12.2.3. Promedio de coeficientes de equivalencia	---	---

2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Varios estudios con muestras grandes	5
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa o KR-20	
2.12.3.3. Promedio de coeficientes de consistencia	Buena-Excelente	4,5

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado y pequeño	3,5
2.12.4.2. Coeficientes de estabilidad	Excelente	5

2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	---	---
2.12.5.2. Coeficientes proporcionados	---	
2.12.5.3. Tamaño de los coeficientes	---	---

2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	---	
2.12.6.2. Promedio de los coeficientes	---	---

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

El manual ofrece, a partir del análisis de las muestras estadounidense y española, diversas evidencias de fiabilidad: consistencia interna, fiabilidad test-retest o estabilidad temporal de las puntuaciones y error típico de medida.

- Consistencia interna: Los datos sobre la consistencia interna de la prueba se basan en varios estudios con varias muestras (se ofrecen datos de consistencia interna para cada edad y para cada versión -entrevista o cuestionario-, oscilando el tamaño muestral de 123 a 258 casos si se tiene en cuenta la combinación de población general y clínica). El alfa de Cronbach promedio se sitúa en torno a .85, por lo que puede considerarse excelente. No obstante, ha de tenerse en cuenta que estos valores se han obtenido combinando las dos muestras (clínica y general) y el número elevado de ítems en cada subescala, lo que contribuye a que sean tan altos. Debe notarse, además, que en población general (no en la muestra combinada) tres subescalas presentan coeficientes inferiores a .70 (Anexo C).
- Estabilidad temporal: las evidencias proporcionadas se consideran excelentes, con coeficientes promedio que oscilan entre .86-.92 en el formato de entrevista y .85-.94 en el cuestionario, calculados sobre una muestra de tamaño moderado (N=183).
- Error típico de medida: resulta destacable que se incluyan los errores típicos de medida para que pueda obtenerse con facilidad el intervalo de confianza de la puntuación verdadera. Estos oscilan de 2.13 a 8.69 puntos en la versión de entrevista y de 1.21 a 7.44 en la versión cuestionario. En el Anexo C se indican los errores máximos de medida para el cálculo de los intervalos de confianza, resultando más elevados en población clínica de muy corta edad y en población general a partir de los 10 años, algo que se advierte en el Manual. Esta circunstancia conlleva que deban interpretarse con cautela las puntuaciones obtenidas por estos grupos.

En general, parece que los índices de fiabilidad son algo mejores para la versión cuestionario.

Puesto que en la descripción del manual se enfatiza la utilidad de la prueba en la evaluación de programas e incluso se ejemplifica la misma con un caso en el que los profesores adecúan sus contenidos curriculares a las debilidades registradas por sus alumnos en el DP-3, sería interesante realizar un estudio sobre fiabilidad interjueces (profesionales vs. progenitores).

Asimismo, y habiendo utilizado TRI en la calibración de los ítems de la prueba, podría haberse incluido información sobre la Función de Información de cada uno de los ítems del test, sobre todo teniendo en cuenta que se trata de una prueba de 'screening' de posibles retrasos en el desarrollo y que, por tanto, interesa medir con más precisión en la zona de habilidad comprendida en torno a la media y dos desviaciones típicas por debajo de ésta.

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	Amplio rango de baremos en función de la edad	5
2.13.1.2. Tamaño de las muestras	Muy grande	5
2.13.1.3. Aplicación de estrategia de tipificación continua	Sí	
2.12.1.4. Procedimiento de selección de las muestras	Aleatorio, con muestras representativas de la población objetivo	
2.12.1.5. Actualización de baremos	Excelente	5

2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	---	---
2.13.2.2. Procedimiento empleado para fijar los puntos de corte	---	
2.13.2.3. Procedimiento de obtención del acuerdo inter-jueces	---	
2.13.2.4. Valor del coeficiente de acuerdo inter-jueces	---	---

2.13.3. Comentarios generales sobre baremación e interpretación de las puntuaciones:

Los baremos se calculan a partir de una amplia muestra de tipificación (N=1.758), obtenida a partir de un muestreo estratificado por cuotas de la población infantil con edades comprendidas entre 0 y 12 años y 11 meses), atendiendo a características recogidas en el Censo de Población y Viviendas. Además, se delimita en 100 el número de personas a valorar en cada edad, aumentado este número a 200 en las edades de 0, 1 y 2 años al existir mayor variabilidad en lo que a desarrollo de habilidades se refiere en los primeros años de vida. El procedimiento de selección de la muestra clínica (N=469) no se especifica con tanto detalle, existiendo cierta sobrerepresentación de trastornos del neurodesarrollo (TDAH, TEA y DI) frente a otro tipo de condiciones que también son objeto de cribado en el test (problemas motores).

El manual no contiene los baremos, por lo que no pueden ser directamente consultados. La corrección está informatizada, obligándose a utilizar la plataforma de corrección online (Teacorrige.com). El sistema de corrección ofrece 29 baremos para distintos momentos del desarrollo comprendidos entre los 0 y los 12 años y 11 meses. El número de baremos disponibles por año es mayor para los niños de menor edad, en los que se producen más cambios en su desarrollo. El servicio de corrección transforma las puntuaciones directas en escala CI (M=100; DT=15), así como en varias escalas típicas derivadas y centiles, para facilitar la interpretación de la puntuación obtenida. Las puntuaciones tipificadas permiten conocer si el desarrollo del menor se encuentra por encima o por debajo del punto de corte, ofreciendo puntuaciones de fácil interpretación para cada área (directas, centiles y puntuaciones transformadas a escala CI) así como un índice general de desarrollo.

Además, se ofrecen tablas con los errores máximos de medida a un nivel de confianza del 90 y 95%, de modo que resulta relativamente sencillo calcular el intervalo de puntuaciones en el que se encuentra la puntuación verdadera. Ha de tenerse en cuenta que los errores máximos de medida resultan muy elevados en población clínica de muy corta edad y en población general a partir de los 10 años, como se advierte en el Manual. Esta circunstancia implica que se han de interpretar con cautela los resultados obtenidos por estos grupos.

Las normas de interpretación son descritas con mucho detalle en el manual, aspecto que se valora muy positivamente. Se alude en este a consideraciones importantes sobre la validez de las respuestas y los pasos para interpretar la prueba dependiendo de las necesidades de la persona evaluada, de la familia y de la escuela. Se explica cómo interpretar las puntuaciones obtenidas en el test y cómo realizar análisis comparativos entre las escalas y análisis de los ítems. Se incluyen también sugerencias para usar los resultados para elaborar el programa de intervención y se presentan varios casos ilustrativos de los diferentes pasos de la interpretación.

De forma opcional (comprando unos usos específicos del informe), el sistema de corrección permite obtener dos informes narrativos automáticos: un informe para el profesional, en el que se comentan los resultados en cada escala y se realiza un análisis de los ítems y de las diferencias entre las áreas evaluadas; y un informe para la familia, con un resumen adaptado de los principales resultados y una propuesta individualizada de actividades que se pueden realizar para favorecer el desarrollo de la persona evaluada.