

Evaluación de

Cuestionario de madurez neuropsicológica infantil

∞ CUMANIN-2 ∞

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	CUMANIN-2: Cuestionario de madurez neuropsicológica infantil-2
Autor	J. A. Portellano, R. Mateos, R. Martínez Arias, F. Sánchez-Sánchez
Autor de la adaptación española	--
Variable(s)	Madurez neuropsicológica infantil
Áreas de aplicación	Neuropsicología
SopORTE	Administración oral Papel y lápiz Manipulativo

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	Excelente	5
Fundamentación teórica	Buena	4
Adaptación	--	--
Análisis de ítems	Excelente	5
Validez: contenido	Buena-Excelente	4.5
Validez: relación con otras variables	Buena-Excelente	4.2
Validez: estructura interna	Excelente	5
Validez: análisis del DIF	Excelente	5
Fiabilidad: equivalencia	--	--
Fiabilidad: consistencia interna	Excelente	5
Fiabilidad: estabilidad	Buena-Excelente	4.3
Fiabilidad: TRI	Adecuada	3
Fiabilidad inter-jueces	--	--
Baremos e interpretación de puntuaciones	Excelente	5

Comentarios generales

El CUMANIN-2 es la versión revisada del CUMANIN publicado en 1999. La revisión realizada ha permitido mejorar algunas pruebas e ítems y actualizar los baremos. La fundamentación teórica sobre la que se sustenta es adecuada, aunque para futuras ediciones, los apartados sobre neuroplasticidad, funciones encefálicas, diferenciación hemisférica y trastornos del desarrollo necesitarían ser actualizados y referenciados de manera más adecuada según la literatura internacional en neuropsicología.

Se trata de un test para evaluar la madurez neuropsicológica infantil en niños desde 3 hasta 6 años y 11 meses de edad. El test supone un sistema integrado de exploración neuropsicológica mediante la evaluación de tres funciones cognitivas básicas como son el desarrollo sensoriomotor, memoria y aprendizaje y lenguaje. Permite obtener tres índices parciales: el Índice de Desarrollo Sensoriomotor (ISM), que se obtiene a partir de las pruebas de psicomotricidad, estructuración espacial, visopercepción, ritmo y atención; el Índice de Memoria y Aprendizaje (IMA), derivado de las pruebas de memoria visual y verbal; y el Índice de Desarrollo del Lenguaje (IDL), que se calcula a partir de las pruebas de lenguaje articulatorio, expresivo, comprensivo y de fluidez verbal. Además, es posible obtener una puntuación global como el Índice de Desarrollo Neuropsicológico (IDN) a partir de los tres índices anteriores. El test está compuesto por un total de 14 subescalas, que se dividen en puntuaciones principales y adicionales. Cabe destacar que el test ofrece instrucciones claras para obtener las distintas puntuaciones (directas y transformadas), así como la obtención de los índices global y parciales y la interpretación que se debe hacer de las distintas puntuaciones obtenidas.

La evaluación de los aspectos anteriores proporciona información sobre el grado de desarrollo alcanzado en cada una de las áreas examinadas, lo que permite elaborar un perfil cognitivo y neuropsicológico para identificar los puntos fuertes y áreas de mejora de los niños evaluados. La detección de determinadas áreas de mejora facilita la planificación y el diseño de intervenciones, cuya eficacia puede ser evaluada a partir de criterios explícitos presentes en el manual.

A grandes rasgos, se trata de un test sustentado en una teoría sólida y que es ampliamente utilizado en la evaluación del desarrollo y madurez neuropsicológica de niños entre 3 y 6 años. En este sentido, las evidencias empíricas acumuladas, desde su primera versión, relativas a la validez son extensas y adecuadas. Todas las evidencias de validez aportadas son pertinentes. El estudio de validez de contenido ha empleado una amplia muestra de expertos y se ha llevado a cabo mediante procedimientos sistemáticos, si bien no se indica en el manual el método usado para evaluar el acuerdo entre los jueces. El estudio que valora la evidencia de validez basada en el contenido, lo complementa un estudio piloto que permitió mejorar algunos aspectos relativos a la administración de la prueba.

Un aspecto que tiene muchas repercusiones para la utilidad y consecuencias del uso del test, no siempre analizado, y que, sin embargo, sí ha sido investigado en este test, es el relativo a la capacidad diagnóstica de la prueba. En relación con ello, el análisis de curvas ROC, ha demostrado la utilidad, especialmente del IDN para detectar sujetos con necesidad de atención más específica en el área del desarrollo neuropsicológico, no obstante, en los casos dudosos la evaluación se debería acompañar de otro tipo de juicio más allá del actuarial.

La estructura interna se corresponde con el modelo teórico usado en el desarrollo de la prueba y su estudio fue llevado a cabo mediante procedimientos adecuados y ajustados a las características de los datos, lo que ha evitado sesgos metodológicos en la interpretación de los resultados. Al estudio de la estructura interna, se le suma un tipo de evidencia muy relevante como es el de la invarianza por sexo.

A grandes rasgos, la valoración general sobre las evidencias de validez aportadas es elevada. Todas las evidencias de validez presentadas en el manual avalan y apoyan el uso pretendido del test. Además, se dan datos muy útiles para poder interpretar las puntuaciones obtenidas.

El análisis de la precisión del test ha sido extenso y profuso, y está extensamente justificada la buena precisión de las distintas medidas ofrecidas por el test.

Finalmente, el proceso de baremación se la llevado a cabo mediante uno de los métodos más potentes para el establecimiento de normas, como es la tipificación continua, y sobre la base de una muestra amplia de más de 1200 participantes. Se ofrecen baremos en intervalos de edad de 4 meses, lo que permite tomar una "foto precisa" del estado de madurez neurocognitiva de los niños evaluados y, además, posibilita valorar su desarrollo. Los perfiles obtenidos a partir de las puntuaciones en las distintas pruebas facilitan la detección de puntos débiles en los sujetos evaluados, aspecto que se hace esencial para la planificación de intervenciones eficaces y eficientes desde el punto de vista de la evaluación de programas. El informe de corrección es adecuado y útil para informar a las familias.

En términos generales, resulta un test bien desarrollado, cuyas evidencias empíricas apoyan y avalan el uso pretendido del mismo.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

CUMANIN-2: Cuestionario de madurez neuropsicológica infantil-2

1.2. Nombre del test en su versión original:

--

1.3. Autor/es del test original:

J.A. Portellano, R. Mateos, R. Martínez-Arias, F. Sánchez-Sánchez

1.4. Autor/es de la adaptación española:

--

1.5. Editor del test en su versión original:

Hogrefe TEA Ediciones

1.6. Editor de la adaptación española:

--

1.7. Fecha de publicación del test original:

2021

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

--

1.9. Fecha de la última revisión del test:

2021

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

Neuropsicología

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

El test consiste en un sistema integrado de exploración neuropsicológica que evalúa tres funciones cognitivas fundamentales: desarrollo sensoriomotor, memoria y aprendizaje y lenguaje. Está compuesto por un total de 14 pruebas, de las cuales 11 son principales y 3 adicionales. Estas pruebas abarcan diversos dominios cognitivos, como lenguaje articulatorio, lenguaje comprensivo, lenguaje expresivo, visopercepción, psicomotricidad, estructuración espacial, ritmo, memoria verbal, memoria visual, lateralidad, atención, fluidez verbal, lectura y escritura.

Las áreas de evaluación se dividen en:

- a) Procesamiento sensoriomotor, que incluye las pruebas 1. Psicomotricidad, 4. Estructuración espacial, 6. Visopercepción, 8. Ritmo y 12. Atención.
- b) Memoria y aprendizaje, que abarca las pruebas 2. Memoria visual y 3. Memoria verbal.
- c) Lenguaje, que se evalúa mediante 4 pruebas: 5. Lenguaje articulatorio, 7. Lenguaje expresivo, 9. Lenguaje comprensivo y 11. Fluidez verbal.

Además, se incluyen tareas adicionales que constan de 3 pruebas: 10. Lateralidad, 13. Lectura y 14. Escritura, estas dos últimas aplicables a partir de los 5 años de edad.

A partir de las diferentes pruebas, es posible obtener tres índices globales: el Índice de Desarrollo Sensoriomotor (ISM) (compuesto por las pruebas de psicomotricidad, estructuración espacial, visopercepción, ritmo y atención), el Índice de Memoria y Aprendizaje (IMA) (que incluye las pruebas de memoria visual y memoria verbal) y el Índice de Desarrollo del Lenguaje (ILE) (compuesto por las pruebas de lenguaje articulatorio, expresivo, comprensivo y fluidez verbal). Además, a partir de los tres índices anteriores se puede calcular el Índice de Desarrollo Neuropsicológico (IDN), el cual permite resumir en una única puntuación el grado de madurez neuropsicológica del niño o niña. A partir de las 14 pruebas es posible calcular:

- Puntuaciones principales: son las que se combinan para obtener cada uno de los tres índices (ISM, IMA e ILE) que deberían ser la base de la interpretación de los resultados.

- Puntuaciones adicionales: proporcionan información complementaria sobre alguna de las áreas evaluadas (lectura y escritura en el área de lenguaje), o aportan información transversal a todas las áreas (lateralidad).

La evaluación de los aspectos anteriores proporciona información sobre el grado de desarrollo alcanzado en cada una de las áreas examinadas, lo que permite elaborar un perfil cognitivo y neuropsicológico para identificar los puntos fuertes y áreas de mejora de los niños evaluados. La detección de determinadas áreas de mejora facilita la planificación y el diseño de intervenciones, cuya eficacia puede ser evaluada a partir de criterios explícitos presentes en el manual.

La duración total de administración es de 50 minutos aproximadamente, el formato de aplicación es en papel y lápiz y el modo de corrección online. Permite determinar el nivel de Desarrollo madurativo, identificar posibles trastornos del neurodesarrollo u otras dificultades, identificar casos con necesidades educativas especiales, así como obtener el perfil de los puntos fuertes y débiles de cada niño, entre otros aspectos.

Resulta de especial interés el correlato que se establece entre las funciones cognitivas evaluadas por cada prueba y las bases neuroanatómicas implicadas.

1.12. Áreas de aplicación:

Neuropsicología

1.13. Formato de los ítems:

Respuesta dicotómica (sí/no, verdadero/falso, etc.). En algunas pruebas, como por ejemplo las de memoria o fluidez, se obtiene una escala cuantitativa definida por el número de estímulos que recuerda de un total de 9, en el caso de las pruebas de memoria, o por el número de palabras de las oraciones

construidas en fluidez.

1.14. Número de ítems:

El número total de ítems de la prueba es de 146 (sin contar los 100 estímulos de Atención) que se distribuyen de la siguiente manera en las diferentes pruebas.

Pruebas principales

Psicomotricidad: 16 ítems

Estructuración espacial: 13 ítems

Visopercepción: 14 ítems

Ritmo: 10 ítems

Atención: 100 ítems (100 figuras/estímulos que el participante debe tachar si se trata de un cuadrado)

Memoria visual: 9 ítems

Memoria verbal: 9 ítems

Lenguaje articulatorio: 20 ítems

Lenguaje expresivo: 10 ítems

Lenguaje comprensivo: 10 ítems

Fluidez verbal: 4 ítems

Pruebas adicionales

Lateralidad: 5 ítems (teniendo en cuenta que 10 ítems son comunes a otras pruebas)

Lectura: 13 ítems

Escritura: 13 ítems

1.15. Soporte:

Administración oral, Papel y lápiz, Manipulativo.

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

Nivel C

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

Especifique el rango de edad, nivel educativo, etc. y si el test es aplicable en ciertas poblaciones específicas: minorías étnicas, personas con discapacidad, grupos clínicos, etc.

El test está dirigido a población general en el rango de edad comprendido entre los 3 y los 6 años y 11 meses. Esta edad se corresponde con la etapa de Educación Infantil y el inicio de la Educación Primaria. Sin embargo, no está recomendada, especialmente, en los siguientes casos:

Alteraciones graves del lenguaje.

Limitaciones auditivas, visuales y motoras importantes.

Discapacidad intelectual grave.

Para la actualización de la prueba, y poder evaluar el poder discriminativo de la misma, ésta también ha sido administrada a población clínica heterogénea que incluía participantes con discapacidad intelectual, trastornos del espectro autista, TDAH, trastornos de la conducta y emocionales, trastornos del lenguaje o de la comunicación, trastornos motores, nacimiento prematuro y otros problemas del neurodesarrollo. En cualquier caso, la aplicación de la prueba está restringida para los tres casos descritos anteriormente.

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

Algunas pruebas requieren que, para su aplicación, el niño o la niña tenga una determinada edad. Concretamente, las pruebas 13 y 14 se aplican solo a partir de los 5 años.

1.19. Procedimiento de corrección:

Automatizado por ordenador (plataforma de corrección on-line)

1.20. Puntuaciones:

El test destaca por la descripción que realiza para la obtención de las distintas puntuaciones (directas y transformadas), así como en la de sus índices más globales y en la interpretación que se debe hacer de las distintas puntuaciones obtenidas.

El CUMANIN-2 es una prueba que ofrece diferentes tipos de puntuaciones para evaluar el desarrollo neuropsicológico de los niños y niñas. Estas puntuaciones se dividen en tres categorías principales: índices (parciales y global), puntuaciones principales y puntuaciones adicionales.

El Índice General de Desarrollo Neuropsicológico (IDN) se obtiene a partir de los índices parciales. Permite evaluar el rendimiento promedio de la persona evaluada en el conjunto de las tareas. A mayor puntuación en el IDN, mejor es el rendimiento del niño o niña evaluado.

Además del IDN, se ofrecen tres índices parciales, el Índice de Desarrollo Sensomotor (ISM), el Índice de Memoria y Aprendizaje (IMA), y el índice de Desarrollo del Lenguaje (ILE). Estos índices parciales brindan un resumen del nivel de madurez en diferentes áreas.

Las puntuaciones principales son las que se combinan para obtener los tres índices parciales (ISM, IMA e ILE). Estas puntuaciones principales se corresponden con la ejecución de los individuos en las pruebas descritas en el apartado 1.14, y son las que permiten obtener un perfil completo del niño o niña, lo que facilita establecer puntos fuertes y débiles en su desarrollo neuropsicológico. Finalmente, las puntuaciones adicionales proporcionan información complementaria sobre alguna de las áreas evaluadas y/o datos transversales a la evaluación.

Además, la prueba permite realizar una interpretación prueba a prueba, lo que facilita obtener un perfil de los puntos fuertes y áreas de mejora de cada niño o niña.

La prueba ofrece tanto puntuaciones directas como transformadas. Las puntuaciones directas se corresponden directamente con la ejecución del niño o niña en cada prueba y se obtienen a partir del número de ítems correctos, de palabras recordadas o de figuras marcadas por la persona evaluada. Por ello, estas no deben ser interpretadas directamente ni comparadas con el resto de las puntuaciones directas de otras pruebas. Las puntuaciones transformadas que ofrece el CUMANIN-2 se basan en la escala T de McCall (muy habitual en los test comerciales). Además, y resulta muy útil, en el manual se encuentra una tabla que permite realizar directamente la transformación de puntuaciones directas a puntuaciones T. Además, se ofrece un criterio de interpretación basado en percentiles y se establecen puntos de corte mediante curvas ROC con una sensibilidad del 76%.

El manual también considera el error de medida y proporciona intervalos de confianza alrededor de las puntuaciones obtenidas para ayudar al profesional en la toma de decisiones. Se recomienda tener precaución con las puntuaciones que se sitúan en los límites de estos intervalos.

Otro aspecto que resulta de especial relevancia con relación a la interpretación de las puntuaciones es el relativo a la significación de las diferencias. El manual explica, basándose en el método propuesto por Davis (1959), cómo comparar e interpretar las puntuaciones individuales en relación con el promedio del rendimiento obtenido en el resto de las pruebas. El objetivo de este tipo de comparación es planificar intervenciones sobre aquellos aspectos del desarrollo neuropsicológico que se encuentren deficitarios en relación al nivel medio del individuo. Este aspecto está calibrado y en el manual se pueden encontrar los valores críticos para cada prueba que indican las diferencias significativas.

1.21. Escalas utilizadas:

Centiles, Eneatipos, Decatipos, T (Media 50 y desviación típica 10), D (Media 50 y desviación típica 20)

1.22. Posibilidad de obtener informes automatizados:

Sí. El informe representa un perfil neuropsicológico, con puntos fuertes y áreas de mejora de cada niño o niña. Incluye también las curvas de aprendizaje. Es un perfil claro, fácil de interpretar por otros profesionales y padres.

1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: 50 minutos.

En aplicación colectiva: No aplica.

1.24. Documentación aportada por el editor:

Manual

1.25. Precio de un juego completo de la prueba:

180,18 euros (IVA incluido) por un juego completo. Fecha de consulta: 05/09/2023.

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

114,38 euros (IVA incluido) por un Kit de corrección (25 Cuadernillos de anotación + 25 Ejemplares + Pin 25 usos de corrección online). Fecha de consulta: 05/09/2023.

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

Incluidos en el Kit de corrección. Fecha de consulta: 05/09/2023.

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

Incluidos en el Kit de corrección. Fecha de consulta: 05/09/2023.

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	4
2.4. Adaptación del test	--	--
2.5. Desarrollo de los ítems del test	Excelente	5
2.6. Calidad de las instrucciones para el participante	Excelente	5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Excelente	5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Buena	4
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 mediante un procedimiento sistematizado (N > 30)	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	Diferencias entre grupos	
2.11.2.1.2. Tamaño de las muestras	Varios estudios con muestras grandes	5
2.11.2.1.3. Procedimiento de selección de las muestras	<p>Hay dos tipos de muestra: a) la muestra usada para la tipificación de la prueba; y b) muestra clínica.</p> <p>Para la muestra de tipificación se aplica un muestreo estratificado por cuotas en el que estaban representados la edad, el sexo y la región geográfica de la población de 3 a 6 años según el Censo de Población de Viviendas del INE. Se estableció un número de 280 casos para cada grupo de edad, en rangos de 1 año, lo que resultó en un total de 1120 participantes para la muestra final. Cabe añadir que los profesionales que se encargaron de la aplicación en cada región se distribuyeron según entorno (rural, intermedio, urbano) y el tipo de centro educativo (privado, público o concertado). Dentro de cada centro se aleatorizaron los casos y se creó una lista de reserva. Todo lo anterior garantiza la representatividad de la muestra seleccionada.</p> <p>Respecto a la muestra clínica, se indica que los participantes fueron reclutados de centros de atención temprana, de gabinetes o clínicas privadas y de otros centros sanitarios pediátricos especializados que colaboraron en el proyecto. A los profesionales se les facilitó unos criterios de inclusión y exclusión que debían aplicar para seleccionar los casos. La muestra incluye una</p>	

	amplia variedad de trastornos.	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	--	--
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	--	--
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	--	--
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	--	--
2.11.2.1.8. Resultados de las diferencias intergrupo	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	<p>En el caso del CUMANIN-2 no se han utilizado otros test o medidas para obtener evidencias acerca de la validez referida a un criterio. En este caso, y tal y como se indica en las instrucciones del CET-R, el estudio de las evidencias sobre la validez referida a un criterio se ha basado en el análisis de curvas ROC de cada prueba para discriminar la muestra clínica de la muestra general. En este sentido, todas las pruebas han mostrado una buena o muy buena capacidad para discriminar entre ambas muestras, excepto tres pruebas que presentan un comportamiento regular para ello (valores ROC entre 0.68 y 0.85). Concretamente son las pruebas de Memoria Visual, Visopercepción y Atención.</p>	
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 de tamaño moderado o con alguna muestra grande y otras pequeñas	4
2.11.2.2.5. Procedimiento de selección de las muestras	<p>Incidental. Las muestras clínicas provenían de 17 centros diferentes, repartidos en 8 comunidades autónomas, a partir de unos criterios de inclusión/exclusión que se dieron a los profesionales vinculados al reclutamiento. La muestra de comparación se seleccionó extrayendo aleatoriamente de la muestra de tipificación un grupo de casos apareados en</p>	

	edad y sexo con cada uno de los casos de la muestra clínica.	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Adecuada ($0.35 \leq r < 0.45$)-Valores AUC entre 0.68 y 0.85	3

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	Excelente	5

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:

Se trata de un test sustentado en una teoría sólida y que, a partir de su primera versión, es ampliamente utilizado en la evaluación del desarrollo y madurez neuropsicológica de niños entre 3 y 6 años. En este sentido, las evidencias empíricas acumuladas relativas a su validez son extensas, pertinentes y adecuadas.

En lo relativo a la evidencia de validez basada en el contenido, en la actualización del test, se ha llevado a cabo un profundo estudio para la obtención de evidencias empíricas sobre la relevancia y representatividad de los elementos del test. En este sentido, se han revisado otros instrumentos de evaluación neuropsicológica similares, escalas de desarrollo y test de aptitudes. Además, se ha realizado una consulta a 42 expertos de diferentes especialidades para la valoración, entre otros aspectos, de la relevancia y claridad de los ítems, instrucciones, procedimientos de aplicación, o posibles sesgos. Esta consulta se ha llevado a cabo mediante un procedimiento sistematizado. En el manual, se presenta un resumen general de los resultados de la validación de contenido basado en frecuencias, lo que puede ser información suficiente, sin embargo, haber hecho explícito si se consideró algún índice de acuerdo o congruencia entre jueces relativo a la evidencia de validez basada en el contenido, le hubiera dado un valor añadido a esta sección. No obstante, en términos generales, se ha cuidado uno de los aspectos más relevantes en la construcción de un test que es la obtención de evidencias sobre validación de contenido.

Respecto a la evidencia de validez basada en la relación con otras variables, se valoraron las diferencias entre grupos en función del sexo y de la edad. En el caso de la edad se diferenciaron 12 niveles (en grupos de 4 meses desde los 3 a los 6 años) con el objetivo de captar diferencias en las tendencias de desarrollo. Estos resultados indicaron que todas las pruebas se mostraron sensibles a las tendencias de desarrollo en la población general (con desarrollo normotípico), proporcionando evidencias sobre la capacidad de las puntuaciones de las pruebas para mostrar los cambios esperados en cada una de las habilidades que se pretende medir.

Resulta muy relevante el estudio que se realiza de la capacidad predictiva de las pruebas y de los índices parciales y global mediante curvas ROC. Se trata de una evidencia de validez que avala el uso pretendido del test y permite una mejor interpretación de los resultados al ofrecer valoraciones sobre puntos de corte. Los resultados son, en general, satisfactorios y mejoran para los índices parciales y global. A pesar del buen comportamiento general de las pruebas, tres de ellas presentan un AUC que puede considerarse con un rendimiento regular (Memoria visual, Visopercepción y Atención). No obstante, dado el uso pretendido del test, serán las puntuaciones en el IDN las que sirvan para emitir un juicio global sobre el grado de madurez de los niños evaluados y en este sentido, la prueba presenta una sensibilidad del 76%, una especificidad del 82% y una precisión del 79%, valores todos ellos aceptables.

Esto implica que aproximadamente 4 de cada 5 personas evaluadas serán correctamente clasificadas utilizando la puntuación del IDN. No obstante, como en cualquier proceso de evaluación psicológica, es necesario que se complemente la evaluación actuarial con el juicio clínico, particularmente en los casos dudosos.

Con relación al estudio de la estructura interna del test, muy acertadamente, se ha partido de un estudio de las correlaciones entre las pruebas, los tres índices parciales (ISM, IMA e ILE) y el índice global (IDN). Para ello, los autores reportan la matriz de correlaciones de Pearson y ofrecen la significación de estas obtenidas con una muestra de 1.139 participantes (muestra general), por un lado, y de 301 (muestra clínica), por otro. La principal conclusión que se ofrece a partir de estos resultados es que los diferentes aspectos evaluados por las pruebas están relacionados, pero no son redundantes.

Siguiendo con la estructura interna del test, también ejecutan distintos análisis factoriales, tanto desde la perspectiva exploratoria como confirmatoria en muestras distintas. El procedimiento utilizado fue adecuado dadas las características de los datos. Los resultados obtenidos contribuyen a sustentar las distintas puntuaciones en las pruebas, así como los índices parciales y el global. En primer lugar, se evaluó la unidimensionalidad de las pruebas. Una vez confirmado este aspecto para cada una de ellas, a continuación, se realizó un AFE a partir de las puntuaciones de las pruebas, las cuales se agruparon en tres factores relacionados cuyos contenidos se correspondían con memoria y aprendizaje, desarrollo sensoriomotor y lenguaje. Para terminar, mediante un AFC se pusieron a prueba dos modelos, uno de primer orden con tres factores relacionados, y otro de segundo orden en el que el factor de orden jerárquico superior se correspondía con el IDN. Los dos modelos presentaron un buen ajuste. Así que, con el fin de obtener índices parciales para cada área y una puntuación total como indicador del nivel de desarrollo neuropsicológico, se optó por el segundo modelo. Por lo tanto, el buen ajuste del modelo de tres factores de primer orden y un factor de segundo orden apoya la obtención de una puntuación general o total denominada IDN. La elección de los índices globales de ajuste fue adecuada, sin embargo, otros índices globales, como el NNFI o AGFI, hubieran complementado esta información. Dentro del AFC, y como una evidencia de validez muy relevante, se llevó a cabo un análisis multigrupo para evaluar la invarianza factorial entre niños y niñas.

Otra evidencia sobre la validez de la prueba es la relativa al posible funcionamiento diferencial de los ítems en función del sexo. En esta versión se ha realizado un estudio adecuado de esta posible fuente de sesgo sistemático y a pesar de que se han detectado 4 ítems con indicios de DIF, estos no presentaron un tamaño del efecto considerable como para afectar significativamente a la medida de las pruebas.

A modo de resumen, habría que destacar que todas las evidencias de validez presentadas, las comentadas en esta sección como otras no comentadas, avalan y apoyan el uso pretendido del test. Además, se dan datos muy útiles para poder interpretar las puntuaciones obtenidas. Por todo ello, la valoración general en este aspecto es altamente satisfactoria.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad (para cada escala o subescala) Coeficientes de fiabilidad para diferentes grupos de personas Error típico de medida para diferentes grupos de personas Cuantificación del error mediante TRI (Función de información u otros)	

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, Alfa ordinal, Omega, Función de información, Fiabilidad marginal	
2.12.3.3. Promedio de coeficientes de consistencia	Excelente ($r \geq 0.85$)	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 o con alguna muestra grande y otras pequeñas	4
2.12.4.2. Coeficientes de estabilidad	Buena ($0.75 \leq r < 0.80$)- Excelente ($r \geq 0.80$)	4.5

2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	Un estudio con una muestra grande	3
2.12.5.2. Coeficientes proporcionados	Función de Información, Fiabilidad marginal	
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 estudio de la fiabilidad del CUMANIN-2 se aborda desde distintos enfoques. Concretamente, se basa en el estudio de la consistencia interna, estabilidad temporal, función de información y, además, se añade información que permite valorar hasta dónde el cambio observado en una persona tras una intervención se debe a errores aleatorios a la propia práctica.

Con respecto a la consistencia interna, más allá del alfa de Cronbach, se calcula el coeficiente omega así como el alfa y omega ordinal, además de la fiabilidad marginal. En el caso del alfa de Cronbach, los coeficientes reportados están por encima de 0.70, y lógicamente este valor se incrementa para el resto de coeficientes reportados según los otros métodos de estimación de la consistencia interna. El estudio se hace por grupos de edad, sexo y en diferentes muestras (de población general, clínica y combinadas). Además se indican los errores típicos de medida y sus intervalos de confianza, lo que permite contextualizar aún más el dato de la fiabilidad de las escalas.

En cuanto a la estabilidad temporal, además de la correlación corregida entre aplicaciones, se aportan las medias y desviaciones típicas para cada aplicación, un dato que viene a apuntalar los buenos resultados que se obtienen en este sentido. Para facilitar la interpretación también se hace una valoración del tamaño del efecto basado en la *d* de Cohen.

Como comentario general, habría que decir que el estudio de la fiabilidad del test es muy completo, se ha llevado a cabo mediante procedimientos adecuados y que, en términos generales, la precisión de las puntuaciones de las escalas es muy buena.

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, el sexo, el nivel cultural y otras características relevantes	5
2.13.1.2. Tamaño de las muestras	Muy grande ($N \geq 1000$)	5
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	<p>El procedimiento usado para la obtención de la muestra de tipificación está más detallado que el usado para la obtención de la muestra clínica. En el primer caso, se indica que se pretende obtener una muestra representativa de la población española, y para ello, se usa un muestreo estratificado por cuotas en el que estaban representados la edad, el sexo y la región geográfica de la población de 3 a 6 años según el Censo de Población de Viviendas del INE. Se indica que se fijó en 280 el número de casos para cada grupo de edad, en rangos de 1 año, de tal forma que el tamaño de la muestra final quedó fijado, según dicho criterio, en 1120 participantes.</p> <p>Los profesionales que se encargaron de la aplicación en cada región se distribuyeron según entorno (rural, intermedio, urbano) y el tipo de centro educativo (privado, público o concertado). Dentro de cada centro se aleatorizaron los casos y se creó una lista de reserva.</p> <p>Respecto a la muestra clínica, se indica que los participantes fueron reclutados de centros de atención temprana, de gabinetes o clínicas privadas y de otros centros sanitarios pediátricos especializados que colaboraron en el proyecto. A los profesionales se les facilitó unos criterios de inclusión y exclusión que debían aplicar para seleccionar los casos. La muestra incluye una amplia variedad de trastornos.</p> <p>Por lo tanto, la muestra general está aleatorizada, las muestras clínicas son intencional.</p>	
2.12.1.5. Actualización de baremos	Excelente (menos de 10 años)	5

2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	Excelente	5
2.13.2.2. Procedimiento empleado para fijar los puntos de corte	Análisis de curvas ROC	
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:

En el manual se explica que el proceso de baremación se llevó a cabo mediante tipificación continua. Aunque no están presentes en el manual, hay baremos establecidos para 12 grupos de edad, lo que implica segmentar la muestra en intervalos de 4 meses. En los baremos, se ofrecen las correspondencias entre los percentiles con las escalas T, S, decatipos y enatipos.

El proceso de baremación se ha basado en una muestra amplia de más de 1200 participantes reclutados en diferentes regiones de España y tratando de que cada grupo de edad, dividido en años de 3 a 6 años, estuviese bien representado. Aunque la muestra se ha segmentado en tramos de 4 meses para el establecimiento de los baremos, el hecho de haber aplicado tipificación continua garantiza resultados robustos.

El perfil resumen facilita la interpretación de las puntuaciones. El punto de corte es la T de 40, que tiene buena sensibilidad y especificidad para detectar aquellos niños con más riesgo y, por tanto, susceptibles de evaluación más amplia o intervención.

Cabe destacar que, se añade el cálculo del índice de cambio fiable y su significación a diferentes niveles de significación (i.e., $p < .10$; $p < .05$; $p < .01$), con el fin de facilitar la interpretación clínica de los cambios de puntuaciones cuando se llevan diferentes evaluaciones en diferentes periodos temporales.