

EVALUACIÓN DEL TEST

WISC-IV

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Características	Descripción
Nombre del test	Escala de inteligencia de Weschler para niños (WISC-IV)
Autor	David Weschler
Autor de la adaptación española	Sara Corral, David Arribas, Pablo Santamaría, Manuel J. Sueiro y Jaime Pereña
Editor de la adaptación española	departamento I+D, TEA Ediciones S.A.
Fecha de la última revisión del test en su adaptación española	2005
Constructo evaluado	Inteligencia
Área de aplicación	Todas en las que se evalúe la inteligencia en niños y adolescentes
Soporte	Administración oral Manipulativo; lápiz y papel

Valoración del test

La escala utilizada en las valoraciones tiene un rango comprendido entre 1 y 5. El 1 indica una valoración baja y el 5 excelente.

Características	Valoración	Puntuación
Calidad de los Materiales y documentación	Excelente	5
Fundamentación teórica	Excelente	4,5
Adaptación Española	Excelente	4,5
Análisis de los ítems	Excelente	5
Validez de contenido	Excelente	5
Validez de constructo	Adecuada	3
Análisis del sesgo	No se aporta	-
Validez predictiva	Bueno	4
Fiabilidad: equivalencia	No se aporta	-
Fiabilidad: consistencia interna	Buena	4
Fiabilidad: estabilidad	Adecuada	3,5
Baremos	Buena	4

Comentarios generales

El test de WISC-IV es una **excelente prueba** para la exploración de las capacidades cognitivas. La tarea realizada por TEA para el logro de su rápida adaptación a España ha sido encomiable, siendo probablemente el test de inteligencia más utilizado a nivel mundial. En España constituye claramente un test de primera elección para valorar la inteligencia (Muñiz y Fernández-Hermida, 2010). Por otra parte no existe en España ninguna otra batería de inteligencia de estas características (K-ABC-2, Stanford Binet 5ª edic, Woodcock Johnson-III, DAS, etc), el cuestionario goza de excelentes propiedades psicométricas, es fácil de aplicar y el material es atractivo para el niño. La corrección es en general sencilla, tiene una buena adaptación a la población española y es un test de reciente publicación y tipificación

Entre los **puntos fuertes de la prueba** cabe citar su mayor sustento conceptual en línea con los actuales estudios jerárquicos de la inteligencia (**modelo de Cattell-Horn-Carroll**), introduciendo el análisis de la inteligencia cristalizada, el de la inteligencia fluida, el procesamiento visual, y el procesamiento de información, especialmente de las habilidades de Memoria de Trabajo y de Velocidad de Procesamiento. Así, la versión actual se organiza en torno a **cuatro grandes índices que representan habilidades intelectuales generales** (Comprensión verbal y Razonamiento perceptivo) y habilidades de procesamiento cognoscitivo (Memoria de Trabajo y Velocidad del Procesamiento). Esta organización de habilidades excluye la clásica y dicotómica clasificación de la inteligencia en CIV y CIM, y también excluye la identificación del tercer factor o Distractibilidad, que queda subsumido entre las habilidades de Memoria de Trabajo y de Velocidad de procesamiento. En los índices de Comprensión Verbal y de Razonamiento Perceptivo se ha eliminado tanto la influencia de la velocidad de respuesta como el énfasis concedido a los aprendizajes. El Razonamiento Perceptivo se diferencia del anterior CIM en que ahora se centra más en razonamiento fluido y procesamiento visual. El énfasis otorgado a la velocidad como indicador cognoscitivo queda relegado a las pruebas que conforman el índice de Velocidad de Procesamiento y a la prueba de Cubos.

Otro de los avances del test es que además de permitir el análisis de las habilidades en relación a los grupos normativos, en función de la edad, **ofrece pautas muy claras para analizar las habilidades en términos intraindividuales**. En este sentido, permite contrastar los distintos tipos de índices entre sí, bien sea en relación al referente muestral de edad, o bien al nivel de aptitud alcanzado, y también permite identificar las debilidades y fortalezas específicas de un sujeto, con distintas posibilidades comparativas (en función de la media global del sujeto, o de las medias específicas de Comprensión verbal y de Razonamiento perceptivo, y a distintos niveles de confianza).

Entre los **puntos débiles** cabe destacar los siguientes:

- El test, en esta versión, **es menos sensible a las dificultades de aprendizaje**: las pruebas de Información y Aritmética, han quedado relegadas a pruebas optativas, que completan la interpretación de los índices de Comprensión verbal y Memoria de Trabajo, respectivamente, pero que no los configuran. **Los clínicos echan en falta las pruebas de Historietas y Rompecabezas**, que permitían analizar la inteligencia visuo-espacial y favorecían la observación de respuestas de perseveración y reacciones frente a la frustración. Se ha criticado que **la Memoria de trabajo sólo incluya tareas auditivas** (números, letras y problemas aritméticos) y que no incluya tareas de tipo no verbal, tales como span de golpes (como la prueba del xilófono del McCarthy), o memoria espacial (como en el K-ABC). También se ha criticado que la **prueba de Conceptos no satura adecuadamente en ninguno de los índices**, aunque la versión española ofrece, en general, datos muy consistentes al respecto. Finalmente destacar la crítica de que **el test es excesivamente redundante**, en el sentido de que hay demasiadas pruebas centradas en el análisis de la inteligencia general, algunas muy cuantitativas, y que se han marginado otros aspectos del funcionamiento intelectual. A nivel interpretativo la diferencia entre CIV y CIM permitía observar las capacidades no verbales en grupos bilingües, grupos de inmigrantes y en grupos desfavorecidos. No es posible ahora realizar esta interpretación. A este mismo nivel interpretativo las pruebas de Comprensión y las de Historietas permitían una exploración de la competencia social. Este análisis ha quedado limitado al análisis de la competencia social expresada verbalmente
- Las **mejoras** que pueden ser introducidas en la prueba son **de tipo material** (Datos de madera, cuaderno de elementos plastificado y de mayor consistencia) y **de tipo técnico**. Entre las de tipo técnico cabe citar el incremento de estudios relativos a poblaciones específicas españolas. El manual ofrece muchos estudios sobre los datos técnicos del test en población original pero cuesta a veces diferenciar entre estudios de otras poblaciones y los estudios referidos a España. En este sentido hubiera sido necesaria mayor claridad en el Manual técnico y de Interpretación para no dar pie a confusiones.

A modo de conclusión, si un test no precisa presentación entre los psicólogos es el conjunto de pruebas de Wechsler. Las escalas de medida de la inteligencia de Wechsler y de su equipo, cuentan entre las más conocidas, divulgadas mundialmente y empleadas en contextos múltiples. Cabe reconocer su excelente trayectoria histórica anclada en los inicios de la medida psicológica y las continuas adaptaciones realizadas, así como las extensiones a diferentes edades, países, idiomas y culturas. El WISC-IV es un test recomendable, sin duda el más recomendable en su segmento en diversas áreas profesionales.

En Psicología Clínica es un instrumento imprescindible en el diagnóstico del retraso mental (discapacidad intelectual), trastornos de aprendizaje, trastornos del espectro autista, trastorno de déficit de atención con hiperactividad, etc. Así como en numerosos casos de diagnóstico diferencial. Es también un test necesario en cualquier supuesto de evaluación neuropsicológica (traumatismos craneoencefálicos, infecciones del sistema nervioso central, déficits neurocognitivos tras el tratamiento de un tumor cerebral o una leucemia linfoblástica aguda etc). También es de uso habitual en el contexto educativo, en cuanto que permite conocer la capacidad intelectual global, así como capacidades generales y un perfil específico, y es un instrumento importante para basar una intervención educativa. Finalmente también es imprescindible en el contexto forense dado su reconocido prestigio y su excelente validez y fiabilidad.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test

Escala de inteligencia de Wechsler para niños-IV (WISC-IV)

1.2. Nombre del test en su versión original

Wechsler Intelligence Scale for Children, Fourth Edition. Administration and Scoring manual (WISC-IV)

1.3. Autor

David Wechsler y actualmente Psychological Corporation

1.4. Autor de la adaptación española

Sara Corral, David Arribas, Pablo Santamaría, Manuel J. Sueiro y Jaime Pereña.

1.5. Editor del test en su versión original

NCS Pearson, Inc. (Pearson Assessment)

1.6. Editor de la adaptación española

Departamento I+D, TEA Ediciones S.A.

1.7. Fecha de publicación del test original

WISC: 1949

WISC-R: 1974

WISC-III: 1991

WISC-IV: 2003

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

WISC: 1974

WISC-R: 1993

WISC-III: (no adaptado)

WISC-IV: 2005

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

2005

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

Inteligencia

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

El test de WISC-IV para la evaluación de las capacidades intelectuales, en la versión actual, consta de 15 pruebas que se organizan en cuatro índices (CV, RP, MT y VP) y en un CI total. El constructo de inteligencia subyacente a la prueba defiende que las capacidades cognoscitivas se organizan de forma jerárquica, con aptitudes específicas vinculadas a distintos ámbitos cognoscitivos. El WISC-IV ha organizado estos ámbitos en estrecha relación con las actuales teorías de la inteligencia de razonamiento fluido y cristalizado y de Memoria de Trabajo.

- El índice de **Comprensión verbal (CV)** expresa habilidades de formación de conceptos verbales, expresión de relaciones entre conceptos, riqueza y precisión en la definición de vocablos, comprensión social, juicio práctico, conocimientos adquiridos y agilidad e intuición verbal. Costa de cinco pruebas: **Semejanzas (S)** analiza la capacidad de abstraer y generalizar a partir de dos conceptos dados; **Vocabulario (V)** analiza el conocimiento léxico, la precisión conceptual y la capacidad expresiva verbal; **Comprensión (C)** mide razonamiento y juicio social frente a la solución de problemas cotidianos; **Información (I)** evalúa la capacidad de adquirir, conservar y recuperar conocimientos adquiridos y **Adivinanzas (Ad)** mide las habilidades para integrar información, generar conceptos alternativos y condensar información. Las dos últimas pruebas son optativas o no-necesarias para la obtención del índice.
- El índice de **Razonamiento Perceptivo (RP)** expresa habilidades prácticas constructivas, formación y clasificación de conceptos no-verbales, análisis visual y procesamiento simultáneo. Costa de cuatro pruebas: **Cubos (CC)** mide habilidades de análisis, síntesis y organización viso-espacial, a tiempo controlado; **Conceptos (Co)**

mide la formación de conceptos y categorías a partir de material visual; **Matrices (M)** analiza razonamiento por analogías visuales e implica integración de información visual; **Figuras incompletas (FI)**, prueba optativa, analiza las capacidades de reconocimiento y organización perceptiva a tiempo controlado.

- El índice de **Memoria de Trabajo (MT)** analiza la capacidad de retención y almacenamiento de información, de operar mentalmente con esta información, transformarla y generar nueva información. Este índice consta de tres pruebas: **Dígitos (D)** analiza memoria inmediata y memoria de trabajo, indicando habilidades de secuenciación, planificación, alerta y flexibilidad cognitiva; **Letras y Números (LN)** analiza la capacidad de retener y combinar dos tipos de información, organizarla y elaborar un conjunto organizado según consignas; **Aritmética (A)**, que es optativa y con control de tiempo, analiza habilidades de razonamiento numérico, agilidad en el manejo y reorganización de la información, atención y memoria a corto término.
- El índice de **Velocidad de Procesamiento de la información (VP)** mide la capacidad para focalizar la atención, explorar, ordenar y/o discriminar información visual con rapidez y eficacia. Consta de tres subtests que se desarrollan bajo control de tiempo: **Claves (CL)** y **Búsqueda de Símbolos (BS)** miden habilidades de rapidez asociativa, aprendizaje, percepción visual, coordinación viso-manual, atención, motivación y resistencia frente a tareas repetitivas. **Animales (An)** es optativa y analiza atención selectiva, y planificación en la búsqueda ordenada versus desordenada de información.

1.12. Áreas de aplicación

Todas en las que se evalúe la inteligencia en niños y adolescentes

1.13. Formato de los ítems

- **Respuesta libre:** Semejanzas, Vocabulario, Comprensión, Información
- **Elección múltiple**
- **Presentación de modelos a imitar** (praxias constructivas): Cubos
- **Presentación estímulos**
 - a repetir, a reorganizar: Dígitos, Letras y Números
 - a asociar (dibujos o números): Conceptos, Claves, Matrices
 - a completar: Figuras incompletas
- **Tachado de símbolos:** Animales
- **Problemas de resolver:** Aritmética, Adivinanzas

- **Identificación de presencia ausencia de estímulos gráficos:** Búsqueda de Símbolos.

1.14. Número de ítems

Subtests	Ítems de eliminación	Ítems de Presentación/entrenamiento	Ítems efectivos
Cubos		-	14
Semejanzas		1	23
Dígitos			
Orden directo		-	8
Orden inverso		1	8
Conceptos		2	28
Claves			
Claves A (6 y 7 años)		5	59
Claves B (8 a 16 años)		7	119
Vocabulario			36
Letras y números	2	2	(10 x 3) = 30
Matrices		3	35
Comprensión		-	21
Busca de símbolos			
Parte A (6 y 7 años)		4	45
Parte B (8 a 16 años)		4	60
Figuras incompletas		1	38
Animales			
Presentación		37	-
Azar		-	316
Ordenados		-	320
Información		-	33
Aritmética		-	34
Adivinanzas		2	24

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

El tests es aplicable a niños y adolescentes de edades comprendidas entre 6 años 0 meses y 16 años 11 meses. Abarca los niveles educativos de Primaria (6 a 11), Secundaria (12-15) y Bachillerato (primer curso, 16 años)

El test es aplicable preferentemente a grupos de sujetos de los que convenga determinar su nivel de habilidades cognitivas o su funcionamiento neuropsicológico: Diagnóstico de altas capacidades; Diagnóstico de retraso mental leve o moderado; Diagnóstico de Trastornos de aprendizaje; Disfunciones neuropsicológicas. También aplica a poblaciones con alteraciones atencionales, como el TADH y puede ser usado para el diagnóstico de minorías étnicas. Los índices de razonamiento perceptivo y de velocidad de procesamiento, son especialmente adecuados para ello.

Sin embargo los resultados del test pueden infravalorar las capacidades del niño en los casos de discapacidades motoras o de lenguaje, hipoacusia, limitaciones visuales, niños procedentes de otros contextos culturales o con conocimiento limitado del español

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

No existen formas paralelas ni tampoco versiones abreviadas específicas, aunque el test contempla una aplicación con 10 pruebas, y la aplicación completa con 15 pruebas.

1.19. Procedimiento de corrección

- Manual mediante plantilla
- Otros

Algunos subtests se corrigen por plantilla (Claves y Búsqueda de Símbolos), en otros se puntúa la respuesta correcta, de manera sencilla (Conceptos, Matrices, Letras y Números, Dígitos, Aritmética, Figuras Incompletas). En otros subtests la puntuación es

más compleja y requiere la cuidadosa lectura de los criterios y los ejemplos que aparecen en el Manual (Semejanzas, Vocabulario, Comprensión). Pueden consultarse detalladamente la corrección en el Manual de aplicación y corrección

1.20. Procedimiento de obtención de las puntuaciones directas

Existen diferentes sistemas de obtención de la puntuación directa:

- En general se obtiene una puntuación directa a partir de la suma de los aciertos en cada uno de los ítems. Siguen este procedimiento las pruebas de: Conceptos, Claves B, Matrices, Figuras incompletas, Información, Aritmética y Adivinanzas.
- Se dan además las siguientes variaciones:
 - En la prueba de Cubos se puntúa acierto/error, pero se pondera el acierto en la respuesta, según el intento en que se logra el éxito (ítems 1 a 3), la dificultad del ítem (4 puntos para los ítems 4 a 8), y según la rapidez de la respuesta (hasta 7 puntos para ítems 9 a 14). La puntuación directa equivale a la suma de puntos obtenidos a cada ítem.
 - En las pruebas de Semejanzas, Vocabulario y Comprensión se pondera la respuesta en función de la precisión o calidad de la respuesta, en una escala Likert de anclajes en 0 = erróneo y 2 = respuesta acertada precisa. La puntuación directa equivale a la suma de puntos obtenidos a cada ítem.
 - En las pruebas de Dígitos y Letras-Números se puntúa el total de aciertos habidos en cada una de las presentaciones de cada ítem (dos y tres presentaciones para cada ítem respectivamente). La puntuación directa equivale a la suma de puntos dados a cada ítem.
 - En la prueba de Claves A se puntúa como acierto los elementos bien resueltos y se bonifica con puntos la rapidez de la respuesta (si la solución se ha obtenido en menos de 120”). La puntuación directa equivale a la suma de puntos acertados más los puntos de bonificación por defecto de tiempo.
 - En la prueba de Búsqueda de Símbolos la puntuación directa se obtiene calculando la diferencia entre el número de aciertos y el de errores, del total de ítems resueltos.
 - En las pruebas de Animales se obtiene la puntuación directa calculando la diferencia entre aciertos y errores, y bonificando con puntos la rapidez en la respuesta.

1.21. Transformación de las puntuaciones

- Normalizada

1.22. Escalas utilizadas

- Centiles
- Puntuaciones típicas

Se manejan dos tipos de de puntuaciones típicas: puntuaciones escalar ($M = 10$, $DT = 3$) en el caso de las 15 pruebas y puntuaciones Cociente Intelectual (CI) ($M = 100$, $DT = 15$) en términos de puntuaciones compuestas para los cinco índices.

1.23. Posibilidad de obtener informes automatizados

Sí. Se proporciona un informe de 9 páginas que presenta los resultados del test y las comparaciones entre puntuaciones propuestas tanto por el Manual Técnico y de interpretación como por Flanagan y Kaufman (2006). Utiliza términos muy técnicos y no incluye información relevante ajena al test: nivel educativo de los padres, rendimiento escolar, motivo de consulta, observaciones de la conducta, etc., (que sí se incluyen en el WISC-IV writer™) Se advierte en las primeras líneas del informe que *“la interpretación de estos resultados debe ser realizada por un profesional de la Psicología debidamente cualificado”*. El informe automatizado consta de diversas partes:

- **Portada de identificación del test** y del sujeto a quien se refiere (datos personales y de la aplicación).
- **Introducción:** en ella se describen el objeto del informe, el contenido y las instrucciones básicas de interpretación.
- **Resumen de datos.** Incluye cinco tablas:
 - Tabla para el cálculo de la edad.
 - Cuadro de conversión de puntuaciones directas (de las 15 pruebas) a escalares
 - Perfil de puntuaciones escalares de las 15 pruebas según índices
 - Tabla para la conversión de las sumas de escalares según índices en puntuaciones compuestas, percentiles y según intervalos de confianza
 - Perfil de puntuaciones compuestas para los cinco índices

- **Análisis entre e intra-índices y en las habilidades específicas.** Incluye tres formas de análisis:
 - Comparaciones de la significación de las diferencias entre Índices, en función del valor crítico. Comparaciones de la significación de las diferencias entre pruebas y según valor crítico.
 - Análisis intraindividual de puntos fuertes y débiles en las distintas pruebas del tests: comparación de las diferencias entre las pruebas y el total individual (o según habilidad de comprensión verbal, y de razonamiento perceptivo), en función de los valores críticos correspondientes.
 - Análisis del procesamiento de la información: incluye análisis de la influencia del tiempo, o del azar, o de la longitud (numérica) de los ítems en la solución de tareas.
- **Comentarios.** Contiene una explicación de los objetivos del análisis expuesto en la página anterior.
- **Interpretación.** En estas páginas se expone un resumen de las puntuaciones según índices y pruebas, dando el valor del índice en términos de categorías diagnósticas (alto, bajo, normal, dentro de los límites, etc). Sigue una explicación interpretativa de la capacidad intelectual global y una interpretación de los cuatro índices.
- **Comparaciones clínicas.** Aporta comparaciones adicionales en función de las teorías cognoscitivas actuales sobre distintos factores cognitivos. Incluye un resumen de las puntuaciones, y los equivalentes datos en percentiles y según categoría diagnóstica.
- **Observaciones.** Cuando convenga.

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

- Sí

1.25. Tiempo estimado para la aplicación del test

- La ficha técnica del test indica un tiempo de aplicación aproximado de 60 a 75 minutos

1.26. Documentación aportada por el editor

- Manual Técnico y de interpretación

- Libros o artículos complementarios. El editor recomienda el uso del libro: Flanagan, D.P. y Kaufman, A.S.(2006). Claves para la Evaluación con el WISC-IV, que se vende conjuntamente con el material del test
- Otros:
 - Caja con dados bicolores
 - Manual de Aplicación y de corrección
 - Cuaderno de elementos
 - Programa Ayuda a la Corrección e Interpretación, para la instalación de la corrección automatizada.

1.27. Precio de un juego completo de la prueba

- 912 Euros.

1.28. Precio y número de ejemplares del paquete de cuadernillos

- Paquete de cuadernillos de la prueba de Animales (25 ejemplares): 75 Euros

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

- Paquete de cuadernillos de Anotación de respuestas (25 ejemplares): 75 Euros

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

- Servicio de corrección e informe: 19,37 Euros
- Pin con 10 informes: 129,77 Euros
- De 3 a 5 pin, y por pin: 108,14 Euros
- De 6 a 10 pin, y por pin: 86,51 Euros
- Más de 10 pin, y por pin: 64,88 Euros

1.21. Bibliografía básica acerca del test aportada en la documentación

Extensa y seleccionada. El WISC-IV ha generado abundante bibliografía posterior, tanto en la investigación como guías de interpretación. Entre estas últimas hay que destacar:

- Flanagan, D. P. y Kaufman, A. S. (2006). *Claves para la evaluación con el WISC-IV*. Madrid. TEA Ediciones (Ed. orig: *Essentials of WISC-IV Assessment*. John Wiley & Sons. 2004)
- Prifitera, A., Saklofske, D. H. y Weiss, L. G. (Eds.) (2008). *WISC-IV: Clinical Assessment and Intervention*. 2ª edic. Academic Press.
- Sattler, J. M. (2004). *Assessment of Children: WISC-IV and WPPSI-III Supplement J*. Sattler Publisher.
- Sattler, J. M. (2008) *Assessment of children: cognitive foundations* 5ª edic. J. Sattler Publisher.
- Weiss, L. G., Saklofske, D. H., Prifitera, A. y Holdnack, J. A. (2006). *WISC-IV: Advanced Clinical Interpretation*. Academic Press.

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

Aspectos generales

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

Comentarios sobre aspectos generales

Calidad del los materiales del test

- Los **dados de plástico** incluidos se han despegado en alguna ocasión (se rompen o desencajan). Algunas veces las caras blancas (o rojiblancas) presentan una zona teñida de un rojo leve, que puede distorsionar la performance del test
- Las **láminas del cuaderno de elementos** deberían ser plastificadas: los niños las tocan con los dedos (a veces mojados, o sucios) y se deterioran. Convendría que las láminas estuvieran impresas sobre una cartulina más consistente.
- Las **pruebas de Claves y de Búsqueda de Símbolos**, que tienen una presentación diferente según edad, podrían ser ofrecidas en cuadernillo aparte, (para 6 y 7 años, y para 8 y 16). Al incluirlo todo en el cuadernillo de anotaciones se malgasta dinero y papel.

Adaptación del test

- **Sólo existe traducción y adaptación sólo a lengua castellana.** Esta falta de traducción a los idiomas del Estado genera algunas dificultades y confusiones en la aplicación de la pruebas, por ejemplo, en sujetos de habla catalana, en la prueba de L y N se confunde el número 1, con la letra U. Se debería haber tomado esto en cuenta para el futuro.

Validez de contenido

Contenido	Valoración	Puntuación
2.10. Calidad de la representación del contenido o dominio	Excelente	4,5
2.11. Consultas a expertos	Excelente	5

Comentarios sobre validez de contenido

- **Calidad de la representación del contenido o dominio.** En general buena, pero debe tomarse en cuenta que se ha perdido el análisis de la velocidad psicomotora, y del reconocimiento de figuras. a memoria de trabajo sólo es analizada mediante tareas numéricas; faltaría el análisis no-verbales (recurso a imágenes por ejemplo).
- **Consultas a expertos.** En el capítulo 2 del Manual se expone en detalle el procedimiento seguido en la adaptación española para lograr un contenido válido: se hicieron pruebas pilotos adicionales, que sustentaran la elección de los ítems.

Validez de constructo

Contenido	Valoración	Puntuación
2.12. Diseños empleados	<ul style="list-style-type: none"> ▪ Correlacionales con otros test. ▪ Diferencias entre grupos ▪ Análisis Factorial Exploratorio ▪ Análisis Factorial Confirmatorio ▪ Estudio de grupos especiales; validez cruzada por edades 	
2.13. Tamaño de las muestras	Excelente	5
2.14. Procedimiento de selección de las muestras	<ul style="list-style-type: none"> ▪ Aleatorio 	
2.15. Correlaciones del test con otros test similares	Bueno	4
2.16. Calidad de los test empleados como criterio o marcador	Adecuado	3
2.17. Datos sobre el sesgo de los ítems	No se aporta	

Comentarios sobre la validez de constructo

- **Tamaño de las muestras:** La validación del constructo se ha realizado según grupos de edad y según toda la muestra. Por un lado se utilizó una muestra de 1590 niños para la tipificación general del test, los análisis factoriales y las diferencias entre grupos.
- **Correlación del test con otros similares** (validez convergente). En la adaptación española se han establecido correlaciones entre las pruebas del WISC-IV y el Raven en una muestra de 49 sujetos. Las correlaciones entre pruebas e índices más próximos conceptualmente con el Raven son variables y poco consistentes (entre .17 y .64). En ningún caso la correlación es superior a .64 (Raven y Figuras incompletas). El análisis se llevó a cabo con un grupo de 49 niños españoles.
- **Datos sobre el sesgo de los ítems.** Este análisis no se ha llevado a cabo con datos españoles por diversos motivos. En primer lugar porque un análisis de sesgo basado en TRI requiere acumular una cantidad considerable de sujetos, en torno a 200, que hayan dado respuesta a cada uno de los ítems de cada subtest. Aunque la muestra de tipificación de la adaptación española es lo suficientemente grande, no todas las personas contestaron a todos los ítems, puesto que están ordenados por dificultad y existen reglas de terminación que impiden que, en la mayoría de los casos, no se llegue al final del test. Para hacer un análisis de este tipo se requiere hacer aplicaciones adicionales sin tiempo a una muestra diferente de la de tipificación.

Y puesto que el WISC-IV español es una adaptación del WISC-IV americano y ambas versiones comparten un gran porcentaje de los ítems, aunque ocupen puestos diferentes por su dificultad, no se consideró necesario replicar los análisis de sesgo que sí han sido realizados con la versión original (para más información consúltase la página 11 del Technical and interpretive manual, 2003) y que permitieron llegar a una versión final de cada subtest con los mejores ítems.

Por otro lado, se consideró que, puesto que ya había quedado demostrada la bondad de los ítems en la versión original, resultaba más adecuado estudiar cómo esos ítems funcionaban en la población española mediante los correspondientes análisis de dificultad, discriminación y contribución a la fiabilidad. Los resultados de estos análisis determinaron la ordenación de los ítems de cada subtest así como la eliminación de aquellos que mostraron un mal funcionamiento.

Validez predictiva

Contenido	Valoración	Puntuación
2.18. Diseño de selección del criterio	No se aporta	
2.19. Tamaño de las muestras	Bueno	4
2.20. Procedimiento de selección de las muestras	▪ Incidental	
2.21. Correlaciones del test con los criterios	Buena	3

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	-

Consistencia interna

Contenido	Valoración	Puntuación
2.25. Tamaño de las muestras	Buena	4
2.26. Coeficientes de consistencia	Excelente	4,5

Estabilidad (test-retest)

Contenido	Valoración	Puntuación
2.27. Tamaño de las muestras	Un estudio con una muestra moderada (100<N<200)	2
2.28. Coeficientes de estabilidad	Excelente	5

Comentarios sobre fiabilidad

Los estudios de fiabilidad realizados en la versión original americana tienen resultados similares en los coeficientes de consistencia, sin embargo en la estabilidad, la muestra americana es de 243 niños, mientras que la española es de 57, y la mediana de los coeficientes de estabilidad para los diferentes Índices en la muestra americana está entre 0,85 y 0,93, mientras que en la española oscila entre 0,75 y 0,91.

Baremos

Contenido	Valoración	Puntuación
2.29. Calidad de las normas	Excelente	4,5
2.30. Tamaño de las muestras	Excelente	5
2.31. Procedimiento de selección de las muestras	Aleatorio	

Comentarios sobre baremación

- **Procedimiento de selección de las muestras.** Aleatorio estratificado. A partir del Censo de población para cada Autonomía se fijaron los sujetos por edad, en valores absolutos y según porcentajes por región o categoría creada. Se fijaron cuatro regiones (N, S, E y Centro). Se distribuyeron las aplicaciones por examinadores por región según zona de residencia (urbana, intermedia o rural) y tipo educativa (público, privado o concertado). Cada examinador aplicó entre 40 y 50 pruebas (especificándole sexo y edad de aplicación). Se ha conseguido una buena representación según el censo español en relación a la representación de las regiones. No obstante la representación urbana está sobre ponderada y la distribución del tipo de escuelas según regiones es irregular. Más representación privada en las zonas Norte y este y más escuelas públicas en zonas sur y Centro.
- **En general son adecuados y la valoración es positiva.** No obstante, los baremos para las edades de 6;0-6;3, 6;4-6;7 y 6;8-6;11, en las pruebas de Datos, Semejanzas, Dígitos, y Comprensión, son poco adecuados para la discriminación de niños con dificultades. Faltan ítems más fáciles para estas edades y para los niveles cognoscitivos bajos. En las edades más elevadas (de 14;4 a 16;11) parece que el baremo por grupos de edad (de tres meses en tres meses) es innecesario, dado que la evolución de las adquisiciones es muy escasa o muy lenta. Es decir, **falta mayor sensibilidad evolutiva en algunas pruebas.** Por ejemplo, en la prueba de Dígitos la

puntuación directa de 17 ofrece el mismo valor escalar de 10 para las edades de 14;4 a 16;11. También para el valor escalar de 15 se da la misma puntuación de 23 entre las edades de 14;4 a los 16;3. Algo parecido ocurre en las pruebas de Conceptos y de forma parecida en matrices. Resulta necesario clarificar si este hecho es debido al desarrollo del constructo analizado, o al tipo de ítems empleados.

- **El Manual no ofrece las medidas y las dispersiones por grupos de edad**, ello dificulta una opinión más clara acerca de la calidad de los baremos.
- En la tipificación americana la estratificación se hizo por edad, sexo, raza, nivel educativo de los padres y región geográfica. **En la tipificación española no se ha tenido en cuenta el nivel educativo de los padres.**
- La distribución por regiones geográficas (norte, sur, este, centro), plantea algunas dudas en cuanto a los criterios de atribución a esas zonas. **No se tiene en cuenta el caso de las comunidades bilingües**, en las que el dominio del español de los niños más pequeños no es completo.
- Tanto en lo que se refiere a los baremos americanos como a los españoles se han realizado sobre una muestra grande, mayor de 1500 sujetos, sin embargo hay que tener en cuenta que **cada una de las tablas de baremos**, que cubre un rango de cuatro meses, **está realizada sobre datos de unos 50 niños únicamente.**