PsychCorp

1

WPPSI-IV

ESCALA DE INTELIGENCIA DE WECHSLER
PARA PREESCOLAR Y PRIMARIA-IV

Manual técnico
y de interpretación



ALWAYS LEARNING

PEARSON

PS/WPPSI-IV

WPPSI-IV

ESCALA DE INTELIGENCIA DE WECHSLER
PARA PREESCOLAR Y PRIMARIA-IV

Manual técnico y de interpretación

David Wechsler

UAM
BIBLIOTECA DE PSICOLOGÍA

WPPSI-IV/1

PsychCorp

Alemania Australia/Nueva Zelanda Canadá Dinamarca Estados Unidos Francia Holanda India Reino Unido Suecia/Noruega

Índice

Agradecimientos de Pearson Clinical & Talent Assessment III

Ficha técnica XIII

Prólogo XV

Capítulo 1.
Introducción 1

Corrientes históricas y actuales de la evaluación de la inteligencia 1

Perspectiva histórica y teórica de las escalas de Wechsler 2

Consideraciones sobre la evaluación de la inteligencia en los niños pequeños 4

Capítulo 2.
Diseño y objetivo de la escala 7

Versiones anteriores 7

Contenido y terminología de las pruebas 8

Contenido de las pruebas 8

Cambios generales 8

Pruebas suprimidas 8

Pruebas conservadas 9

Pruebas nuevas 9

Cambios en las pruebas 9

Descripción de las pruebas 10

Pruebas de Comprensión verbal 10

Pruebas de Visoespacial 11

Pruebas de Razonamiento fluido 12

Pruebas de Memoria de trabajo 12

Pruebas de Velocidad de procesamiento 13

Terminología de las pruebas 14

Contenido y denominación de las puntuaciones compuestas y las puntuaciones de procesamiento 14

Puntuaciones compuestas 14

Terminología de las puntuaciones compuestas 15

Puntuaciones de procesamiento 15

Estructura de la WPPSI-IV 15

Edades de 2:6 a 3:11 años 16

Edades de 4:0 a 7:7 años 17

Objetivos de la revisión 18

Actualización de las bases teóricas 18

Modelos estructurales de la inteligencia contemporáneos 19

Investigación sobre el neurodesarrollo y la neurocognición 19

Modelos e investigación sobre la memoria de trabajo. 21
Mayor adecuación al desarrollo 22
Materiales manipulativos. 23
Consignas 23
Pruebas de Velocidad de procesamiento 24
Aumento de la facilidad de uso 24
Seguridad de los ítems 24
Materiales y presentación 25
Tiempo de aplicación. 25
Instrucciones de aplicación y corrección 25
Reglas de terminación 26
Mejora de las propiedades psicométricas 26
Ítems y reglas de puntuación. 26
Baremos. 26
Fiabilidad y validez. 26
Suelo y techo de las pruebas 26
Sesgo de los ítems 27
Opciones del nivel de significación de los valores críticos 27
Incremento de la utilidad clínica. 27
Estructura de la escala 27
Metodología de comparación de las diferencias entre puntuaciones. 28
Ampliación del rango de edad 29
Requisitos de lenguaje expresivo. 29
Puntuaciones de los índices secundarios 29
Grupos especiales 30

Capítulo 3.
Desarrollo de la versión estadounidense de la WPPSI-IV 31
Directrices de investigación. 31
Principales fases de la investigación 31
Desarrollo conceptual. 31
Equipo asesor 31
Investigación profesional 32
Estudios semiestructurados de expertos y examinadores. 32
Estudios piloto. 32
Experimentación 32
Tipificación 33
Composición final y evaluación de la WPPSI-IV 33
Consistencia de los ítems, las consignas y los estímulos. 33
Consistencia del orden de aplicación 33

Capítulo 4.
Adaptación española de la WPPSI-IV 35
Fase experimental 35
Selección de los ítems experimentales. 35

Selección y formación de los examinadores ... 37
Aplicación de la versión experimental ... 37
Control y seguimiento ... 37
Muestra experimental ... 38
Análisis de los datos de la fase experimental ... 38
Fase de tipificación ... 39
Modificaciones respecto a la WPPSI-III ... 39
Diseño de la muestra de tipificación ... 40
Selección y formación de los examinadores ... 41
Aplicación de la versión de tipificación ... 41
Puntuación de los protocolos ... 41
Muestra de tipificación real ... 42
Representatividad de la muestra ... 43
Creación de los baremos ... 46
Establecimiento de los puntos de comienzo y de las reglas de terminación ... 46
Obtención de las puntuaciones escalares de las pruebas ... 46
Obtención de las puntuaciones compuestas ... 47
Cálculo de las puntuaciones de edad equivalente ... 48

Capítulo 5.
Fiabilidad ... 49
Fiabilidad y errores de medida ... 50
Consistencia interna ... 50
Errores típicos de medida e intervalos de confianza ... 54
Estabilidad temporal: test-retest ... 57
Acuerdo entre correctores ... 61
Diferencias entre puntuaciones ... 61
Significación estadística de las diferencias entre puntuaciones ... 61
Valores críticos corregidos y no corregidos ... 62
Tasas base de las diferencias entre puntuaciones ... 64
Análisis de dispersión entre pruebas ... 65
Resumen ... 65

Capítulo 6.
Validez ... 67
Evidencia basada en el contenido ... 67
Evidencia basada en los procesos de respuesta ... 68
Evidencia basada en la estructura interna ... 69
Estudios de intercorrelación ... 70
Intercorrelaciones de pruebas, puntuaciones de procesamiento y puntuaciones compuestas ... 71
Análisis factorial confirmatorio ... 75
Evidencias basadas en las relaciones con otras variables ... 85
Relaciones con otras medidas ... 85
Correlaciones con la WPPSI-III ... 86

Correlaciones con las Bayley-III 88
Correlaciones con la WISC-IV 90
Correlaciones con el NNAT2 93
Correlaciones con la NEPSY-II 94
Correlaciones con el BASC-2 CPP 99
Grupos especiales 101
Niños con alta capacidad intelectual 103
Niños con discapacidad intelectual leve o moderada 105
Niños con retraso del desarrollo cognitivo 109
Niños con factores de riesgo para el desarrollo 111
Niños con problemas previos a la lectoescritura 113
Niños con trastorno por déficit de atención con hiperactividad 115
Niños con comportamiento perturbador 117
Niños con trastorno del lenguaje expresivo 119
Niños con trastorno mixto del lenguaje receptivo-expresivo 121
Niños con trastorno autista 123
Niños con trastorno de Asperger 125
Resumen del rendimiento de los grupos especiales en la WPPSI-IV 127
Consecuencias de la aplicación de tests 127

Capítulo 7.
Interpretación de los resultados 129
Descripción de los resultados 129
Puntuaciones típicas 129
Percentiles 131
Errores típicos de medida e intervalos de confianza 132
Clasificación cualitativa 132
Puntuaciones de edad equivalente 133
Etapas para el análisis de un perfil 133
Realización del análisis primario de un perfil 134
Paso 1. Descripción del CI total 134
Paso 2. Descripción del ICV 134
Paso 3. Descripción del IVE 135
Paso 4. Descripción del IRF (solo edades 4:0-7:7) 135
Paso 5. Descripción del IMT 135
Paso 6. Descripción del IVP (solo edades 4:0-7:7) 135
Paso 7. Evaluación de los puntos fuertes y débiles en los índices 136
Paso 8. Evaluación de las comparaciones entre los índices 137
Paso 9. Evaluación de los puntos fuertes y débiles en las pruebas 138
Paso 10. Evaluación de las comparaciones entre las pruebas 138
Realización del análisis secundario de un perfil 139
Paso 1. Descripción del IAV 139
Paso 2. Descripción del INV 139
Paso 3. Descripción del ICG 139

Paso 4. Descripción del ICC (solo edades 4:0-7:7) 139
Paso 5. Evaluación de las comparaciones entre los índices 140
Paso 6. Evaluación de las comparaciones entre las pruebas de los índices secundarios 140
Paso 7. Evaluación de las comparaciones entre las puntuaciones de procesamiento (solo edades 4:0-7:7) 140
Realización del análisis cualitativo de las respuestas 141
Resumen 141

Anexo A. Tablas de intercorrelaciones 143

Anexo B. Uso del IAV y del INV 153
Introducción 153
Aplicaciones del IAV y del INV 153
Obtención del IAV y del INV 154
Interpretación básica del IAV y del INV 158
Interpretación del IAV 158
Interpretación del INV 159

Anexo C. Uso del ICG y del ICC 161
Descripción 161
Introducción al ICG y al ICC 161
Aplicación del ICG y del ICC 162
Obtención del ICG y del ICC 164
Interpretación básica del ICG y del ICC 168
Evaluación del ICG y del ICC 168
Comparación y evaluación de las diferencias entre el ICG y el CI total y entre el ICG y el ICC 168

Anexo D. Criterios de inclusión para los grupos especiales 173
Criterios de inclusión generales para los grupos especiales (muestra estadounidense) 173
Criterios de inclusión específicos para los grupos especiales (muestra estadounidense) 174
Alta capacidad intelectual 174
Discapacidad intelectual 174
Retraso del desarrollo cognitivo 174
Factores de riesgo para el desarrollo 174
Problemas previos a la lectoescritura 175
Trastorno por déficit de atención con hiperactividad 175
Comportamiento perturbador 175
Trastorno del lenguaje expresivo 175
Trastorno mixto del lenguaje receptivo-expresivo 176
Trastorno autista 176

Trastorno de Asperger. 176

Bibliografía 177

Tablas

Tabla 2.1. Abreviaciones de las pruebas de la WPPSI-IV 8
Tabla 2.2. Cambios en las pruebas de la WPPSI-IV 9
Tabla 2.3. Abreviaciones de las puntuaciones compuestas de la WPPSI-IV 14
Tabla 4.1. Ítems conservados, eliminados y nuevos respecto a la versión estadounidense para la versión experimental española 36
Tabla 4.2. Cambios en la adaptación españolad de la WPPSI-IV respecto a la WPPSI-III. 40
Tabla 4.3. Tamaño de la muestra de tipificación por sexo y por grupo de edad. 42
Tabla 4.4. Porcentajes teóricos (censo) y reales (muestra de tipificación), según edad y nivel educativo 43
Tabla 4.5. Porcentajes teóricos (censo) y reales (muestra de tipificación), según edad y tipo de población 44
Tabla 4.6. Porcentajes teóricos (censo) y reales (muestra de tipificación), según edad y zona geográfica. 44
Tabla 4.7. Porcentajes teóricos (censo) y reales (muestra de tipificación), según edad y sexo . . . 44
Tabla 5.1. Coeficientes de fiabilidad de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas (datos españoles). 51
Tabla 5.2. Coeficientes de fiabilidad de las pruebas para los grupos especiales (datos estadounidenses) 53
Tabla 5.3. Errores típicos de medida de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas (datos españoles) 55
Tabla 5.4. Datos demográficos del estudio de estabilidad test-retest y del estudio de acuerdo entre correctores (datos españoles) 57
Tabla 5.5. Coeficientes de estabilidad de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas (datos españoles) 58
Tabla 6.1. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas, por franja de edad. 72
Tabla 6.2. Índices de bondad de ajuste para el análisis factorial confirmatorio de todas las pruebas. 80
Tabla 6.3. Índices de bondad de ajuste para el análisis factorial confirmatorio de las pruebas de los índices primarios 83
Tabla 6.4. Datos demográficos para los estudios de validez con otras medidas 85
Tabla 6.5. Correlaciones entre la WPPSI-IV y la WPPSI-III (datos españoles) 86
Tabla 6.6. Rangos de las puntuaciones esperadas para la WPPSI-IV correspondientes a determinadas puntuaciones compuestas de la WPPSI-III (datos españoles) 87
Tabla 6.7. Correlaciones entre la WPPSI-IV y las Bayley-III 88
Tabla 6.8. Rangos de las puntuaciones esperadas para la WPPSI-IV correspondientes a determinadas puntuaciones compuestas de las Bayley-III 90
Tabla 6.9. Correlaciones entre la WPPSI-IV y la WISC-IV 91
Tabla 6.10. Rangos de las puntuaciones compuestas esperadas para la WISC-IV correspondientes a determinadas puntuaciones compuestas de la WPPSI-IV 92

Tabla 6.11. Correlaciones entre la WPPSI-IV y el NNAT2 93
Tabla 6.12. Correlaciones entre la WPPSI-IV y la NEPSY-II. 95
Tabla 6.13. Correlaciones entre la WPPSI-IV y el BASC-2 CPP 100
Tabla 6.14. Datos demográficos de los grupos especiales (datos estadounidenses) 102
Tabla 6.15. Resultados del grupo con alta capacidad intelectual y del grupo control (datos estadounidenses) 104
Tabla 6.16. Resultados del grupo con discapacidad intelectual leve y del grupo control (datos estadounidenses) 107
Tabla 6.17. Resultados del grupo con discapacidad intelectual moderada y del grupo control (datos estadounidenses) 108
Tabla 6.18. Resultados del grupo con retraso del desarrollo cognitivo y del grupo control (datos estadounidenses) 110
Tabla 6.19. Resultados del grupo con factores de riesgo para el desarrollo y del grupo control (datos estadounidenses) 112
Tabla 6.20. Resultados del grupo con problemas previos a la lectoescritura y del grupo control (datos estadounidenses) 114
Tabla 6.21. Resultados del grupo con TDAH y del grupo control (datos estadounidenses) 116
Tabla 6.22. Resultados del grupo con comportamiento perturbador y del grupo control (datos estadounidenses) 118
Tabla 6.23. Resultados del grupo con trastorno del lenguaje expresivo y del grupo control 120
Tabla 6.24. Resultados del grupo con trastorno mixto del lenguaje receptivo-expresivo y del grupo control 122
Tabla 6.25. Resultados del grupo con trastorno autista y del grupo control (datos estadounidenses) 124
Tabla 6.26. Resultados del grupo con trastorno de Asperger y del grupo control (datos estadounidenses) 126
Tabla 7.1. Relación entre las puntuaciones escalares, las desviaciones típicas respecto a la media y los percentiles equivalentes. 130
Tabla 7.2. Relación entre las puntuaciones compuestas, las desviaciones típicas respecto a la media y los percentiles equivalentes. 131
Tabla 7.3. Clasificación cualitativa de las puntuaciones compuestas 132
Tabla A.1. Intercorrelaciones de las puntuaciones de las pruebas y de las puntuaciones compuestas para las edades 2:6-2:11 144
Tabla A.2. Intercorrelaciones de las puntuaciones de las pruebas y de las puntuaciones compuestas para las edades 3:0-3:5 144
Tabla A.3. Intercorrelaciones de las puntuaciones de las pruebas y de las puntuaciones compuestas para las edades 3:6-3:11 145
Tabla A.4. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 4:0-4:5. 146
Tabla A.5. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 4:6-4:11 147
Tabla A.6. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 5:0-5:5 148
Tabla A.7. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 5:6-5:11 149

Tabla A.8. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 6:0-6:11 150
Tabla A.9. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 7:0-7:7 151
Tabla B.1. Conversión suma de puntuaciones escalares a IAV para las edades de 2:6 a 7:7 años (datos españoles)..... 155
Tabla B.2. Conversión suma de puntuaciones escalares a INV para las edades de 2:6 a 3:11 años (datos españoles)..... 156
Tabla B.3. Conversión suma de puntuaciones escalares a INV para las edades de 4:0 a 7:7 años (datos españoles)..... 157
Tabla C.1. Conversión suma de puntuaciones escalares a ICG para las edades de 2:6 a 3:11 años (datos españoles)..... 165
Tabla C.2. Conversión suma de puntuaciones escalares a ICG para las edades de 4:0 a 7:7 años (datos españoles)..... 166
Tabla C.3. Conversión suma de puntuaciones escalares a ICC para las edades de 4:0 a 7:7 años (datos españoles)..... 167
Tabla C.4. Valores críticos para las diferencias estadísticamente significativas entre el ICG y determinadas puntuaciones compuestas, por grupo y franja de edad..... 169
Tabla C.5. Porcentajes acumulados (tasas base) de las diferencias entre el ICG y determinadas puntuaciones compuestas, a partir de toda la muestra de tipificación y según nivel de aptitud en el ICG..... 170

Figuras

Figura 2.1. Estructura de la WPPSI-IV para las edades de 2:6 a 3:11 años 16
Figura 2.2. Estructura de la WPPSI-IV para las edades de 4:0 a 7:7 años 17
Figura 4.1. Distribución de los examinadores de la fase de tipificación 43
Figura 4.2. Comparación entre los datos del censo y los datos de la muestra de tipificación según las características demográficas 45
Figura 6.1. Modelo jerárquico de tres factores para todas las pruebas (2:6-3:11 años)..... 81
Figura 6.2. Modelo jerárquico de cinco factores para todas las pruebas (4:0-7:7 años)..... 82
Figura 6.3. Modelo jerárquico de tres factores para las pruebas de los índices primarios (2:6-3:11 años)..... 84
Figura 6.4. Modelo jerárquico de cinco factores para las pruebas de los índices primarios (4:0-7:7 años)..... 84

Ficha técnica

Nombre del test:	*Escala de inteligencia de Wechsler para preescolar y primaria-IV*
Nombre original:	*Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition*
Autor:	David Wechsler
Adaptación española:	Departamento de I+D de Pearson Clinical and Talent Assessment: Ana Hernández, Elena de la Guía, Erica Paradell, Frédérique Vallar
Editor original:	NCS Pearson Inc, 2012
Editor de la adaptación española:	*Pearson Educación, 2014*
Aplicación:	Individual
Ámbito de aplicación:	Niños de edades comprendidas entre 2 años 6 meses y 7 años 7 meses.
Tiempo de aplicación:	Entre 60 y 90 minutos aproximadamente
Objetivo:	Evaluación completa de la aptitud intelectual del niño.
Tipificación:	Muestra representativa de la población española compuesta de 885 niños, teniendo en cuenta los siguientes criterios de estratificación: edad, sexo, nivel educativo, zona geográfica y tipo de población.
Material:	*Manual técnico y de interpretación*, *Manual de aplicación y corrección*, cuadernillo de anotación (edad 2:6-3:11), cuadernillo de anotación (edad 4:0-7:7), cuaderno de estímulos 1-3, cuadernillo de respuestas 1-3, plantilla de corrección para *Búsqueda de animales*, plantilla de corrección para *Cancelación*, plantilla de corrección para *Clave de figuras*, 14 Cubos para *Cubos*, 3 planos de *Localización*, Tarjetas de animales para *Localización*, 13 Puzles para *Rompecabezas*, Tampón.
Nivel de cualificación requerido:	Nivel C; titulación superior en psicología, psiquiatría o psicopedagogía y experiencia profesional en diagnóstico clínico.
Corrección:	Manual o mediante plataforma on-line.

Introducción

La *Escala de inteligencia de Wechsler para preescolar y primaria-IV* (WPPSI-IV) es un instrumento clínico de aplicación individual diseñado para evaluar la inteligencia de los niños de 2 años 6 meses a 7 años 7 meses (2:6 a 7:7 años). Se trata de una versión revisada de la *Escala de inteligencia de Wechsler para preescolar y primaria-III* (WPPSI-III; Wechsler, 2002, 2009). La WPPSI-IV ofrece puntuaciones de los índices primarios, que reflejan el funcionamiento intelectual en determinadas áreas cognitivas (p. ej., Índice de comprensión verbal, Índice visoespacial, Índice de memoria de trabajo); una puntuación compuesta que representa la aptitud intelectual general (CI total); y puntuaciones de los índices secundarios, que reflejan otros aspectos de la aptitud cognitiva (p. ej., Índice no verbal, Índice de capacidad general). Para presentar el contexto de la WPPSI-IV, en este capítulo se describen las corrientes históricas y actuales de la evaluación de la inteligencia y se exponen consideraciones sobre la evaluación de la inteligencia en los niños más pequeños.

Corrientes históricas y actuales de la evaluación de la inteligencia

A principios del siglo xx, la teoría dominante sobre la inteligencia defendía un único constructo de inteligencia general subyacente, en gran parte responsable del rendimiento del sujeto en todas las tareas mentales. Spearman (1904) identificó este constructo como factor *g*. Aunque su teoría de la inteligencia incluía otros factores específicos, los primeros tests de inteligencia pretendían clasificar a los sujetos según su nivel general de funcionamiento cognitivo (Beres, Kaufman y Perlman, 2000; Goldstein y Hersen, 2000; Wasserman, 2012).

En 1905, Binet y Simon, a petición del gobierno francés, publicaron una escala de inteligencia con el objetivo de desarrollar métodos que permitieran identificar a aquellos niños que podrían tener dificultades para seguir una educación normal. Continuando con las mejoras que se habían realizado desde principios del siglo xx, Terman y sus colegas de la Universidad de Stanford (1916) presentaron una versión ampliada y tipificada en Estados Unidos. Aunque la metodología de desarrollo de los tests fue mejorando a lo largo de los años, el principal objetivo de los tests de inteligencia seguía siendo la identificación del déficit intelectual.

La entrada de Estados Unidos en la primera guerra mundial generó la necesidad de crear medidas de inteligencia para evaluar y clasificar a los reclutas. Con este fin se creó el *Army Alpha*, un test que incluía un considerable componente verbal. Pero la alta tasa de analfabetismo de algunos reclutas hizo que fuera necesaria una medida no verbal de inteligencia, por lo que se desarrolló el *Army Beta* (Thorndike, 1997). Wechsler reconoció esta doble necesidad de evaluar la inteligencia de forma verbal y no verbal, y, al elaborar su primera escala de inteligencia, la *Wechsler-Bellevue Intelligence Scale* (Wechsler-Bellevue; Wechsler, 1939), incluyó puntuaciones tanto para la escala verbal como para la escala manipulativa, además de una puntuación total. La Wechsler-Bellevue también era innovadora porque proporcionaba las puntuaciones CI de desviación, basadas en puntuaciones típicas calculadas con las mismas características de distribución para todas las edades.

En la década de 1950, la evaluación de la inteligencia se centró en la medición de aspectos más específicos del funcionamiento cognitivo de los sujetos. Durante este periodo, los avances en las técnicas de análisis factorial se aplicaron al campo de la medición para aclarar aun más la naturaleza de la inteligencia. Cattell, que había sido alumno de Spearman, formuló la teoría de que la inteligencia se componía de dos factores generales: la inteligencia fluida (*Gf*) y la inteligencia cristalizada

UAM BIBLIOTECA DE PSICOLOGÍA

(*Gc*; Cattell, 1941, 1957). Más tarde, Horn amplió la teoría de Cattell e incluyó los siguientes factores: percepción visual, memoria a corto y largo plazo, velocidad de procesamiento, aptitud de procesamiento auditivo, razonamiento cuantitativo, lectura y escritura (Horn, 1985, 1988, 1991; Horn y Noll, 1997). Mientras los investigadores identificaban dominios de inteligencia más específicos, la identificación de los tests de inteligencia comenzó a centrarse en el rendimiento del sujeto en ámbitos más segmentados del funcionamiento cognitivo, así como en las posibles diferencias de desarrollo en esos dominios (Carroll, 1993; Goldstein y Hersen, 2000; Keith, 1985, 1990; Schneider, Schumann-Hengsteler y Sodian, 2005).

A lo largo de los últimos setenta años, gran parte del debate sobre la evaluación de la inteligencia se ha centrado en la existencia de un aspecto global subyacente de la inteligencia, que influye en el rendimiento intelectual del sujeto a través de distintos dominios cognitivos (Gustafsson y Undheim, 1996; Jensen, 1998; Reeve y Charles, 2008). Carroll (1993, 2012), basándose en los estudios realizados hasta la fecha que evaluaban las pruebas de aptitud cognitiva mediante el análisis factorial, concluyó que la evidencia de un factor general de inteligencia era incuestionable. Por tanto, a pesar de la tendencia a considerar múltiples, y más definidas, aptitudes intelectuales, continúa siendo válido el concepto de un factor global subyacente de inteligencia (véase Johnson, Bouchard, Krueger, McGue y Gottesman, 2004).

La mayoría de las teorías sobre la inteligencia actuales consideran que su estructura es jerárquica, con un factor general en el nivel superior y amplios dominios de aptitud cognitiva en los niveles siguientes; estos dominios, a su vez, divididos en aptitudes más específicas y segmentadas. Los resultados basados en los estudios del análisis factorial coinciden en la identificación de entre 8 y 10 grandes dominios de inteligencia (Carroll, 1993, 2012; Horn y Blankson, 2012; Johnson et al., 2004). Es necesario continuar con las investigaciones para determinar si esos dominios son los objetivos de evaluación adecuados para los niños pequeños. En los niños más pequeños se han demostrado empíricamente menos factores generales de inteligencia respecto a los niños más mayores y los adultos (Morgan, Rothlisberg, McIntosh y Hunt, 2009; Ward, Rothlisberg, McIntosh y Bradley, 2011). Sin embargo, cada vez son más numerosas las demostraciones empíricas para una serie de aptitudes diferenciadas en los niños más pequeños; entre ellas, el conocimiento cristalizado, la memoria a corto plazo, la percepción visual, la recuperación a largo plazo y, en algunos casos, el razonamiento fluido (p. ej., Kaufman, Reynolds, Liu, Kaufman y McGrew, 2012; Kuwajima y Sawaguchi, 2010; Morgan et al., 2009; Schneider et al., 2005).

Perspectiva histórica y teórica de las escalas de Wechsler

La historia de la WPPSI-IV se inicia con la Wechsler-Bellevue. Wechsler basó este test en la premisa de que la inteligencia es una entidad global, porque caracteriza el comportamiento del sujeto como un todo, y, al mismo tiempo, una entidad específica, compuesta por aptitudes o elementos distinguibles entre ellos. A partir de su experiencia clínica, Wechsler seleccionó y desarrolló pruebas destinadas a poner de relieve los aspectos cognitivos de la inteligencia que le parecían importantes medir: comprensión verbal, razonamiento abstracto, procesamiento visoespacial, razonamiento cuantitativo, memoria y velocidad de procesamiento. Las teorías y pruebas de inteligencia actuales han confirmado que estas áreas son aspectos importantes de la aptitud cognitiva (Carroll, 1993, 2012; Horn y Blankson, 2012; Schneider y McGrew, 2012).

Algunos investigadores consideraron que Wechsler, dado que divide las pruebas en tareas verbales y manipulativas, asumió una estructura bifactorial de la inteligencia. Sin embargo, Wechsler (1958) expuso el objetivo práctico de tal división:

> [La agrupación de las pruebas en las áreas verbal y manipulativa] no implica que estas sean las únicas aptitudes abordadas en los tests. [...] Las pruebas son diferentes medidas de la inteligencia, no medidas de diferentes tipos de inteligencia; y la dicotomía de áreas verbal y manipulativa es solo una de las muchas formas posibles de clasificar los tests. (p. 64)

En las revisiones más recientes de las escalas de inteligencia de Wechsler, se ha ampliado la medida de dominios específicos del funcionamiento cognitivo (p. ej., memoria de trabajo y velocidad de procesamiento), en atención a las teorías sobre la inteligencia, los estudios de análisis factorial y las investigaciones clínicas actuales, sin dejar de ofrecer una medida fiable de la inteligencia global (CI total).

Aunque la evaluación y división del funcionamiento cognitivo en dominios más específicos presenta ventajas, existen otros aspectos que deben tenerse en cuenta. Primero, las funciones cognitivas están interrelacionadas, funcional y neurológicamente, lo que dificulta la medición de un dominio puro del funcionamiento cognitivo. Incluso las medidas más tradicionales de dominios específicos, como la velocidad de procesamiento, implican que el sujeto tenga la capacidad de comprender la instrucción verbal, discriminar entre los estímulos visuales, procesar la información e indicar la respuesta mediante la intervención del sistema motor. Pese a que los resultados de los estudios de análisis factorial sugieren la presencia o ausencia de los dominios evaluados, pueden no reflejar la variedad de aptitudes cognitivas requeridas para realizar una prueba concreta. Dado que las cargas factoriales de cada prueba difieren según la composición y combinación de las pruebas, la interpretación de lo que mide una prueba concreta varía en función de la combinación de pruebas incluidas en el análisis.

Segundo, es ecológicamente válido incluir pruebas que requieren la participación de varias aptitudes cognitivas para realizar correctamente las tareas. Las tareas cognitivas raramente se llevan a cabo de forma aislada. Como escribió Wechsler (1975):

> [...] los atributos y factores de la inteligencia, al igual que las partículas elementales en la física, tienen al mismo tiempo propiedades individuales y colectivas, es decir, parecen comportarse de forma distinta cuando están solas a como lo hacen cuando funcionan conjuntamente. (p. 138)

La validez ecológica de la inteligencia general se demuestra por su capacidad de predecir cosas tales como el rendimiento laboral (Hunt y Madhyastha, 2012; Kuncel, Ones y Sackett, 2010), la salud mental y física y los hábitos de salud (Johnson, Corley, Starr y Deary, 2011), así como el rendimiento académico y el nivel educativo (Deary y Johnson, 2010; Deary, Strand, Smith y Fernandes, 2007; Johnson, Deary y Iacono, 2009; Kaufman et al., 2012). Los distintos dominios de la aptitud cognitiva, amplios y específicos, que comprenden la inteligencia no muestran el mismo grado de capacidad predictiva (Gottfredson, 2008; Hartmann, Larsen y Nyborg, 2009; Kotz, Watkins y McDermott, 2008; Reeve y Charles, 2008). La evaluación de aspectos concretos mediante factores psicométricamente puros puede ser útil en investigación o para identificar déficits neuropsicológicos específicos, pero ello no se traduce necesariamente en información relevante en términos de la expresión de una conducta inteligente en la vida real (Jacobson, Delis, Hamilton, Bondi y Salmon, 2004).

Tercero, no sería razonable pensar que una única medida de inteligencia pueda evaluar todos los dominios de forma significativa y práctica (Flanagan, Ortiz y Alfonso, 2007). Wechsler logró seleccionar medidas representativas de una amplia variedad de dominios cognitivos (p. ej., comprensión verbal, visoespacial, memoria), de los que se había demostrado su importancia como aspectos del funcionamiento cognitivo. También advirtió que durante una sesión de evaluación, la fatiga del examinador o del sujeto podía producir resultados inválidos, por ello seleccionó un número suficiente de pruebas que proporcionasen información clínicamente significativa en un tiempo de aplicación razonable. Wechsler pensaba que otros factores relacionados, como el rendimiento académico y las habilidades motoras, podían influir en el rendimiento del sujeto en los tests de inteligencia, pero consideraba que estos aspectos se evaluaban mejor con pruebas específicamente elaboradas para estos dominios. Su sensibilidad hacia los aspectos prácticos de la evaluación de la inteligencia se demuestra en el hecho de que elaboró una escala exclusivamente para la evaluación de la memoria (Wechsler, 1945).

Wechsler creía que los resultados de las pruebas de aptitud cognitiva solo reflejaban una parte de la inteligencia, a la que definió como la «capacidad del sujeto para actuar con determinación, para pensar racionalmente y para relacionarse eficazmente con el entorno» (1944). Wechsler era muy consciente de que los resultados de los estudios factoriales representaban solo una parte de la inteligencia, y creía que había otros atributos que contribuían a la inteligencia. Entre estos atributos figuran la planificación y conciencia de los objetivos, el entusiasmo, la dependencia o independencia de campo, la impulsividad, la ansiedad y la perseverancia.

La influencia de Wechsler en el campo de la evaluación de la inteligencia sigue presente en la actualidad. Las recomendaciones actuales para las buenas prácticas indican que, al interpretar los resultados de los tests, los profesionales han de considerar al sujeto como único y tener en

cuenta otros atributos además de la inteligencia (Kaufman y Lichtenberger, 2006; Sattler, 2008a). Comúnmente, se acepta que los sujetos que obtienen puntuaciones similares en los tests pueden afrontar con distinta eficacia los retos planteados por el entorno, debido a aspectos no relacionados con sus aptitudes cognitivas. A la inversa, dado que algunos aspectos que no guardan relación con la inteligencia influyen en los resultados de los tests, los sujetos con distintos niveles de aptitud intelectual pueden obtener puntuaciones similares. La evaluación de la inteligencia de un sujeto es más compleja que la simple obtención de sus puntuaciones en las pruebas de inteligencia. Como indicó Wechsler (1975):

> Lo que medimos con los tests no es lo que los tests miden: información, percepción espacial, capacidad de razonamiento. Esto es solo un medio para llegar a un fin. Lo que miden los tests de inteligencia es algo mucho más importante: la capacidad del sujeto para comprender el mundo que lo rodea, y sus recursos para afrontar los retos que se le presentan. (p. 139)

Dado que las versiones actuales de las escalas de inteligencia de Wechsler miden tanto el funcionamiento intelectual global como los dominios específicos de la aptitud cognitiva, las bases teóricas de dichas escalas son evidentes. La existencia de una base teórica en las escalas de Wechsler se demuestra claramente en el hecho de que las mismas pruebas, o similares, están presentes en otras escalas de inteligencia, y por las elevadas correlaciones entre las escalas de inteligencia de Wechsler y otras pruebas de aptitud cognitiva. Muchas de las pruebas originales (p. ej., *Cubos* y *Vocabulario*) siguen estando presentes, con alguna modificación, en otras medidas actuales de aptitud intelectual que afirman estar basadas en las teorías de la inteligencia (p. ej., Kaufman y Kaufman, 2004; Roid, 2003). La validez de la escala también se demuestra en sus correlaciones con otras medidas de la aptitud cognitiva (p. ej., *Differential Ability Scales-Second Edition* [DAS-II; Elliott, 2007] *Kaufman Assesssment Battery for Children, Second Edition* [Kaufman y Kaufman, 2004]; *Stanford-Binet Intelligence Scales, Fifth Edition* [Roid, 2003]; *Woodcock-Johnson III* [Woodcock, McGrew y Mather, 2001]). Las altas correlaciones que se obtienen entre las escalas de inteligencia de Wechsler y otras pruebas de aptitud intelectual indican que estos instrumentos miden constructos similares.

La utilidad práctica y clínica de las escalas de inteligencia de Wechsler se ve avalada por casi setenta años de investigación en distintos campos y con finalidades muy diversas. Una y otra vez, estas escalas han demostrado su utilidad clínica en la identificación de la discapacidad intelectual y de los trastornos del aprendizaje, la asignación a programas especializados, la intervención clínica y la evaluación neuropsicológica (Beres et al., 2000). Dadas las numerosas demostraciones de su utilidad clínica, parece incuestionable que Wechsler tenía una gran conciencia de la naturaleza de la inteligencia cuando seleccionó y desarrolló las pruebas de sus escalas.

Con esta exposición se ha ofrecido una perspectiva histórica y contemporánea de la teoría sobre la inteligencia en relación con la evaluación de la inteligencia en niños mediante la WPPSI-IV. Para una revisión más completa sobre la evolución de la teoría de la inteligencia, véanse Flanagan y Harrison (2012), Goldstein y Beers (2003) y Sternberg (2000).

Consideraciones sobre la evaluación de la inteligencia en los niños pequeños

Evaluar la inteligencia en los niños más pequeños es, en muchos aspectos, un reto. Las características del niño, del profesional y del test pueden afectar a la validez de los resultados. Las características del niño propias de la edad, su nivel de desarrollo, sus experiencias, su estado mental, así como su conducta durante la sesión de evaluación, pueden suponer obstáculos significativos para la interpretación del test. La experiencia del profesional, su enfoque sobre la aplicación de las pruebas, así como el uso de técnicas de evaluación complementarias son variables importantes en la evaluación de la inteligencia en los niños más pequeños. El test debería ser adaptable al nivel de desarrollo del niño, tener unas cualidades técnicas y un diseño adecuados y demostrar su validez para los usos previstos.

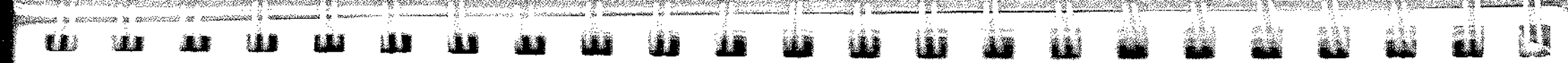
Varios factores relacionados con el niño pueden dificultar la tarea de evaluación de la inteligencia. Es normal que el comportamiento de un niño pequeño varíe cada día y según el contexto en que

se encuentre. Así, un niño puede haber dado anteriormente una determinada respuesta y no volver a darla en otro momento o en un contexto distinto de aplicación de un test (Brassard y Boehm, 2007). El desarrollo cognitivo y físico del niño también puede afectar la aplicación e interpretación del test. Por ejemplo, es posible que el niño no haya adquirido aún las destrezas lingüísticas necesarias para comprender las consignas o para expresar de forma inteligible sus respuestas, o que su desarrollo motor sea insuficiente para realizar correctamente las tareas (Brassard y Boehm, 2007; Ford, Kozey y Negreiros, 2012; Kaufman y Kaufman, 1977; Vig y Sanders, 2007). Las experiencias del niño también pueden influir en la sesión de evaluación y en los resultados. Los niños más pequeños pueden carecer de destrezas para realizar un test, no estar familiarizados con los materiales del test o no haber realizado nunca antes un test (Brassard y Boehm, 2007); de forma que, para ellos, la sesión de evaluación es una situación desconocida en la que no pueden demostrar sus aptitudes completamente. El estado mental del niño y su conducta durante la sesión puede conllevar otras dificultades. Para los niños pequeños siempre es un reto mantener la atención y la concentración, aun en las mejores condiciones; por lo que el cansancio y el malestar emocional pueden hacer que el reto sea significativamente mayor (Ford et al., 2012). El hecho de que durante la aplicación del test no esté con sus padres o cuidadores puede generar malestar emocional o alteración de la conducta en el niño (Brassard y Boehm, 2007).

En la evaluación de la inteligencia de los niños pequeños también son importantes una serie de variables relacionadas con el profesional. Para obtener la información adecuada para la interpretación de los resultados, el profesional ha de tener experiencia en la evaluación de niños pequeños y conocer la literatura relevante sobre la primera infancia, incluidos los modelos de desarrollo cognitivos, motores, lingüísticos y emocionales (Ford et al., 2012). Las respuestas del profesional ante las distintas situaciones que pueden darse durante la aplicación del test también pueden influir en la validez de los resultados. Por ejemplo, el hecho de que el profesional aplique el test siguiendo el ritmo adecuado para el niño o que sepa detectar que el niño necesita un pequeño descanso puede afectar a los resultados (Brassard y Boehm, 2007). Para asegurar la validez de la interpretación, es fundamental disponer de información obtenida mediante otras técnicas de evaluación, como la observación del niño y las entrevistas mantenidas con los padres o cuidadores. Así, el profesional debería incorporar a los resultados del test los datos relevantes sobre el estado actual del niño tanto en su casa como en otros entornos importantes (p. ej., el parvulario) a fin de generar un informe útil y práctico (Brassard y Boehm, 2007; Ford et al., 2012).

Finalmente, las características del propio test también son importantes en la evaluación de la inteligencia. No todos los tests cognitivos presentan la misma sensibilidad al nivel de desarrollo de los niños pequeños (Brassard y Boehm, 2007; Ford et al., 2012). Los tests que incluyen tareas acordes con el desarrollo cognitivo, motor y lingüístico de los niños son mejores, pero un único test no puede satisfacer las necesidades de todos los niños pequeños (Brassard y Boehm, 2007; Ford et al., 2012; Vig y Sanders, 2007). Deberían considerarse distintos aspectos del test, como su rango de edad, materiales, propiedades psicométricas, suelos y techos, dominios cognitivos y tiempo de aplicación. El rango de edad de la WPPSI-IV, que abarca de 2:6 a 7:7 años, puede parecer reducido en comparación con el de otras escalas de inteligencia, pero en este rango se producen cambios cognitivos y del neurodesarrollo fundamentales. Estos cambios suponen retos significativos para los desarrolladores de tests, que deben, dentro de la misma escala, establecer suelos para los niños de 2:6 años, cuyo rendimiento se sitúa en los niveles más bajos, y techos para los niños de 7:7 años, cuyo rendimiento se sitúa en los niveles más altos (Brassard y Boehm, 2007; Ford et al., 2012; Vig y Sanders, 2007). La demostración de la validez clínica del test se basa en estudios con grupos especiales, diseñados para evaluar las diferencias de rendimiento entre dichos grupos especiales y los grupos control de la muestra de tipificación. Otros estudios de validez aportan información sobre los usos prácticos de la escala, como su relación con las medidas de rendimiento académico que permiten evaluar aspectos del aprendizaje durante la primera infancia (Vig y Sanders, 2007).

Queda fuera del alcance de este capítulo exponer con detalle los distintos aspectos que pueden complicar la evaluación de la inteligencia en los niños pequeños. Para una información más extensa al respecto, véanse Brassard y Boehm (2007), Ford et al. (2012), Kaufman y Kaufman (1977) y Vig y Sanders (2007).



Diseño y objetivo de la escala

En los últimos setenta años, las escalas de inteligencia de Wechsler se han revisado frecuentemente para incorporar los avances realizados en el campo de la evaluación de la inteligencia, para actualizar los baremos según los cambios de la población y para adaptarlas a las necesidades prácticas y clínicas de la sociedad contemporánea. El desarrollo de la WPPSI-IV es un paso más en este sentido. En este capítulo se presentan las versiones anteriores de la escala, así como la nueva estructura, las características de las pruebas y el contenido y denominación de las puntuaciones compuestas y de procesamiento. Finalmente, se exponen los objetivos de la revisión.

Versiones anteriores

La primera versión de la *Escala de inteligencia de Wechsler para preescolar y primaria* (WPPSI; Wechsler, 1967, 1976) se creó como una medida de la inteligencia en los niños de 4:0 a 6:6 años para atender a la creciente necesidad de evaluar a la población en edad preescolar. En la siguiente versión, la *Wechsler Preschool and Primary Scale of Intelligence-Revised* (WPPSI-R; Wechsler, 1989; no adaptada a la población española), se conservaron todas las pruebas de la primera versión (*Información*, *Comprensión*, *Aritmética*, *Vocabulario*, *Semejanzas*, *Frases*, *Dibujo geométrico*, *Cuadrados*, *Laberintos*, *Figuras incompletas* y *Casa de los animales*) y se incorporó una prueba nueva: *Rompecabezas*. Además, se amplió el rango de edad, que pasó a ser de 3:0 a 7:3 años, lo cual conllevó que se tuvieran que añadir ítems más fáciles y más difíciles. Los estudios de análisis factorial de ambas escalas apuntaron a un modelo de dos factores de inteligencia: verbal y manipulativo (Carlson y Reynolds, 1981; Gyurke, Stone y Beyer, 1990; Liu y Lynn, 2011; Schneider y Gervais, 1991; Silverstein, 1969; Stone, Gridley y Gyurke, 1991).

Posteriormente se llevó a cabo una revisión más profunda de la escala, que dio lugar a La *Escala de inteligencia de Wechsler para preescolar y primaria-III* (WPPSI-III; Wechsler, 2002, 2009). En esta versión se amplió de nuevo el rango de edad para abarcar a los niños más pequeños, de 2:6 años; y la escala se dividió en dos franjas de edad: de 2:6 a 3:11 años, y de 4:0 a 7:3 años, cada una de ellas con pruebas específicas. De las pruebas de la versión anterior se conservaron siete: *Información*, *Comprensión*, *Vocabulario*, *Semejanzas*, *Cubos* (denominada *Cuadrados* en la WPPSI-R), *Rompecabezas* y *Figuras incompletas*; y se eliminaron cinco: *Frases*, *Dibujo geométrico*, *Laberintos*, *Casa de los animales* y *Aritmética*). Para la WPPSI-III se añadieron siete pruebas nuevas: *Dibujos*, *Nombres*, *Matrices*, *Conceptos*, *Adivinanzas*, *Búsqueda de símbolos* y *Claves*. *Dibujos* y *Nombres* se crearon como pruebas verbales adecuadas para las edades de 2:6 a 3:11 años y como medidas opcionales del vocabulario expresivo y receptivo para las edades de 4:0 a 7:3 años. *Matrices*, *Conceptos* y *Adivinanzas* se incorporaron como medidas del razonamiento fluido para las edades de 4:0 a 7:3 años. Y, finalmente, *Búsqueda de símbolos* y *Claves* se incluyeron como medidas del procesamiento verbal para las edades de 4:0 a 7:3 años. La WPPSI-III ofrecía, para las edades de 2:6 a 3:11 años, tres pruebas verbales (*Información*, *Dibujos* y *Nombres*), dos pruebas manipulativas (*Cubos* y *Rompecabezas*) y cuatro puntuaciones compuestas (CI verbal, CI manipulativo, CI total y Lenguaje general); y, para las edades de 4:0 a 7:3 años, cinco pruebas verbales (*Información*, *Vocabulario*, *Adivinanzas*, *Comprensión* y *Semejanzas*), cinco pruebas manipulativas (*Cubos*, *Matrices*, *Conceptos*, *Figuras incompletas* y *Rompecabezas*), dos pruebas de velocidad de procesamiento (*Claves* y *Búsqueda de símbolos*), dos pruebas opcionales (*Dibujos* y *Nombres*) y cinco puntuaciones compuestas (CI verbal, CI manipulativo, Velocidad de procesamiento, CI

total y Lenguaje general). Las pruebas opcionales permitían obtener la puntuación compuesta Lenguaje general para identificar posibles retrasos del lenguaje, específicamente en el desarrollo del vocabulario.

Contenido y terminología de las pruebas

En esta sección se describe el contenido de las pruebas de la WPPSI-IV y su terminología, así como los cambios realizados en la misma respecto a la WPPSI-III.

Contenido de las pruebas

En la tabla 2.1. figuran las pruebas de la WPPSI-IV, así como su abreviación. A lo largo de este manual, las pruebas se presentan agrupadas por el dominio cognitivo que miden, siguiendo este orden: Comprensión verbal, Visoespacial, Razonamiento fluido, Memoria de trabajo y Velocidad de procesamiento. Dentro de cada dominio cognitivo, las pruebas figuran según el orden de aplicación correspondiente al grupo de edad de 4:0 a 7:7 años.

Tabla 2.1. Abreviaciones de las pruebas de la WPPSI-IV

Prueba principal	Abreviación
Información	I
Semejanzas	S
Vocabulario	V
Comprensión	CO
Dibujos	D
Nombres	N
Cubos	C
Rompecabezas	RO
Matrices	M
Conceptos	CON
Reconocimiento	R
Localización	L
Búsqueda de animales	BA
Cancelación	CA
Clave de figuras	CF

Cambios generales

La WPSSI-IV presenta algunos cambios respecto a la WPPSI-III, tanto en el contenido de las pruebas como en los procedimientos de aplicación y corrección. Es necesario que los examinadores se familiaricen con estos cambios a fin de seguir los criterios estandarizados de aplicación de la escala.

Pruebas suprimidas

En esta versión se eliminaron cuatro pruebas de la WPPSI-III: *Adivinanzas*, *Figuras incompletas*, *Búsqueda de símbolos* y *Claves*. La prueba *Adivinanzas* se eliminó porque su constructo se solapaba con el de *Vocabulario*; sus evidencias de validez como medida del razonamiento fluido no eran sólidas; y su correlación con la prueba *Información* era alta, lo que resultaba redundante psicométricamente. Además, dado que los suelos de *Semejanzas* y *Comprensión* se mejoraron sustancialmente en la WPPSI-IV, la utilidad de *Adivinanzas* en la escala Comprensión verbal se redujo considerablemente. La prueba *Figuras incompletas* se suprimió para posibilitar la inclusión de otras medidas relevantes (p. ej., pruebas de memoria de trabajo). Finalmente, las pruebas

Búsqueda de símbolos y *Claves* se sustituyeron por nuevas medidas de velocidad de procesamiento, más adecuadas al desarrollo de los niños y con menos requisitos de destrezas de la motricidad fina.

Pruebas conservadas

De la WPPSI-III se conservaron cinco pruebas para las edades de 2:6 a 3:11 años: *Información, Dibujos, Nombres, Cubos* y *Rompecabezas*; y diez pruebas para las edades de 4:6 a 7:7 años: *Información, Semejanzas, Vocabulario, Comprensión, Dibujos, Nombres, Cubos, Rompecabezas, Matrices* y *Conceptos*. Aunque se mantuvieron las pruebas, en todas ellas se revisó el contenido de los ítems y las normas de aplicación y corrección.

Pruebas nuevas

Para la WPPSI-IV se crearon cinco pruebas nuevas: *Reconocimiento* y *Localización* para las edades de 2:6 a 7:7 años; y *Búsqueda de animales, Cancelación* y *Clave de figuras* para las edades de 4:0 a 7:7 años. *Reconocimiento* y *Localización* son pruebas de Memoria de trabajo; la primera mide la memoria de trabajo visual y la segunda mide la memoria de trabajo visoespacial. Ambas utilizan la interferencia proactiva en vez de la secuenciación para crear los componentes de procesamiento cognitivo necesarios en las pruebas de memoria de trabajo. Los requisitos de la secuenciación, típicos de las tareas de span que implican ordenación, son incompatibles con el desarrollo de los niños pequeños que no demuestran poseer estrategias de repetición (Bjorklund, Dukes y Brown, 2009; Cowan y Alloway, 2009). *Búsqueda de animales, Cancelación* y *Clave de figuras* son pruebas de Velocidad de procesamiento. *Búsqueda de animales* es una prueba de emparejamiento visual con limitación de tiempo, conceptualmente similar a la prueba *Búsqueda de símbolos* de la WPPSI-III. *Cancelación* es una adaptación de la prueba *Animales* de la *Escala de inteligencia de Wechsler para niños-IV* (WISC-IV; Wechsler, 2003, 2011). *Clave de figuras* es una prueba de pares asociados con limitación de tiempo, conceptualmente similar a la prueba *Claves* de la WPPSI-III. Todas las pruebas nuevas de Velocidad de procesamiento se caracterizan por ser adecuadas al desarrollo de los niños pequeños (p. ej., entretenidas, con dibujos sugerentes, en las que se usa un tampón en vez de un lápiz para indicar las respuestas).

Cambios en las pruebas

La tabla 2.2. muestra los cambios en la WPPSI-IV respecto a la WPPSI-III.

Tabla 2.2. Cambios en las pruebas de la WPPSI-IV

	Prueba nueva	Consignas de aplicación	Registro y puntuación	Ítems nuevos
Información			✓	✓
Semejanzas		✓	✓	✓
Vocabulario			✓	✓
Comprensión		✓	✓	✓
Dibujos				✓
Nombres			✓	✓
Cubos		✓		✓
Rompecabezas				✓
Matrices				✓
Conceptos				✓
Reconocimiento	✓			
Localización	✓			
Búsqueda de animales	✓			
Cancelación	✓			
Clave de figuras	✓			

Descripción de las pruebas

Pruebas de Comprensión verbal

Información

Es una prueba de Comprensión verbal para las edades de 2:6 a 7:7 años. Para los ítems gráficos, el niño debe seleccionar el dibujo que mejor responde a una pregunta sobre conocimientos generales. Para los ítems verbales, el niño debe responder a una serie de preguntas de conocimientos generales. El objetivo de esta prueba es medir la aptitud del niño para adquirir, almacenar y recuperar los conocimientos referidos a hechos generales, pero también se relaciona con la inteligencia cristalizada, la memoria a largo plazo y la aptitud para almacenar y recuperar la información del entorno en que se encuentra y/o de la escuela. Otras habilidades que intervienen en esta prueba son la percepción, la comprensión y la expresión verbales (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008b).

Esta prueba consta de 29 ítems: 4 son ítems gráficos y 25 son ítems verbales. Los 4 ítems gráficos proceden de la WPPSI-III (algunas imágenes fueron ligeramente modificadas). De los 25 ítems verbales, 10 también proceden de la WPPSI-III (algunos de estos con algún pequeño cambio en la formulación) y 15 son nuevos. Para los 14 ítems conservados se revisaron los criterios de puntuación.

Semejanzas

Es una prueba de Comprensión verbal para las edades de 4:0 a 7:7 años. Para los ítems gráficos, el niño observa dos dibujos que pertenecen a una categoría, y debe seleccionar de entre las opciones de respuesta la que pertenece a la misma categoría. Para los ítems verbales, el examinador lee dos palabras que representan objetos o conceptos comunes, y el niño debe describir en qué se parecen. El objetivo de esta prueba es medir el razonamiento verbal y la formación de conceptos pero también se relaciona con la inteligencia cristalizada, la comprensión auditiva, la memoria, el pensamiento asociativo y categórico, la capacidad de distinguir entre características esenciales y secundarias y la expresión verbal (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008b).

Esta prueba consta de 23 ítems: 4 son ítems gráficos y 19 son ítems verbales. Los 4 ítems gráficos son nuevos. De los 19 ítems verbales, 17 proceden de la WPPSI-III y 2 son nuevos. Además, se ha agregado un ítem de ejemplo. Para los 17 ítems conservados se revisaron los criterios de puntuación.

Vocabulario

Es una prueba de Comprensión verbal para las edades de 4:0 a 7:7 años. Para los ítems gráficos, el niño ha de nombrar los dibujos que le muestra el examinador. Para los ítems verbales, el niño ha de definir palabras que el examinador lee. El objetivo de esta prueba es medir el léxico del sujeto, así como la formación de conceptos verbales. Además mide la inteligencia cristalizada, los conocimientos adquiridos, la capacidad de aprendizaje, la expresión verbal, la memoria a largo plazo y el nivel de desarrollo del vocabulario. Otras aptitudes que pueden medirse en esta prueba son la percepción y comprensión auditiva y el pensamiento abstracto (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008b).

Esta prueba consta de 23 ítems: 3 son ítems gráficos y 20 son ítems verbales. De los 3 ítems gráficos, 1 procede de la WPPSI-III y 2 son nuevos. De los 20 ítems verbales, 4 proceden de la WPPSI-III y 16 son nuevos. Para los 5 ítems conservados se revisaron los criterios de puntuación.

Comprensión

Es una prueba de Comprensión verbal para las edades de 4:0 a 7:7 años. Para los ítems gráficos, el niño ha de escoger la opción de respuesta que mejor responde a la pregunta del examinador sobre normas de conducta y reglas sociales. Para los ítems verbales, el niño debe responder a las preguntas que formula el examinador sobre normas de conducta y reglas sociales. El objetivo de esta prueba es medir el razonamiento y conceptualización verbal, la comprensión y expresión verbal, la aptitud para evaluar y utilizar la experiencia, así como para demostrar el conocimiento

y juicio práctico del sujeto. También se relaciona con la inteligencia cristalizada, el conocimiento de las normas de conducta convencionales, el juicio social, la memoria a largo plazo y el sentido común (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008b).

Esta prueba consta de 22 ítems: 4 son ítems gráficos y 18 son ítems verbales. Los 4 ítems gráficos son nuevos. De los 18 ítems verbales, 9 proceden de la WPPSI-III y 13 son nuevos. Para los 9 ítems conservados se revisaron los criterios de puntuación.

Dibujos

Es una prueba de Comprensión verbal para las edades de 2:6 a 7:7 años. El niño debe señalar el dibujo que mejor representa la palabra que lee el examinador. El objetivo de esta prueba es medir el lenguaje receptivo y el desarrollo del vocabulario. También se relaciona con el conocimiento léxico, la información adquirida, la memoria a largo plazo y la percepción de estímulos significativos (Lichtenberger y Kaufman, 2003; Sattler, 2008b).

Esta prueba consta de 31 ítems, 21 de ellos proceden de la WPPSI-III y 10 son nuevos.

Nombres

Es una prueba de Comprensión verbal para las edades de 2:6 a 7:7 años. El niño debe nombrar las imágenes que se muestran en el cuaderno de estímulos. El objetivo de esta prueba es medir el lenguaje expresivo, específicamente en el área del desarrollo semántico (conocimiento léxico). También se relaciona con el conocimiento y la información adquiridos, la memoria a largo plazo y la percepción de estímulos significativos (Lichtenberger y Kaufman, 2003; Sattler, 2008b).

Esta prueba consta de 24 ítems, 17 de ellos proceden de la WPPSI-III y 7 son nuevos.

Pruebas de Visoespacial

Cubos

Es una prueba de la escala Visoespacial para las edades de 2:6 a 7:7 años. El niño debe reproducir, con cubos de uno o dos colores, un modelo construido por el examinador, o bien una imagen mostrada en el cuaderno de estímulos. El niño dispone de un tiempo límite para realizar la tarea. El objetivo de esta prueba es medir la aptitud de analizar y sintetizar estímulos visuales abstractos. También se relaciona con el razonamiento y formación de conceptos no verbales, la inteligencia visual general, la percepción y organización visual, el procesamiento simultáneo, la coordinación visomotora, el aprendizaje y la capacidad de discriminar la figura del fondo en estímulos visuales (Carroll, 1993; Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008b).

Cubos consta de dos partes: parte A y parte B. Para la parte A (ítems 1-8) se utilizan cubos de un solo color y no se penaliza el grado de rotación, aunque sea de 180°. Para la parte B (ítems 9-17) se utilizan cubos de dos colores y se penalizan las rotaciones de 30° o más. Se modificó el procedimiento de aplicación de algunos ítems a fin de facilitar el paso de la presentación del modelo construido por el examinador a la presentación mediante una imagen del cuaderno de estímulos. Al igual que en la WPPSI-III, no se conceden bonificaciones por tiempo dado que los niños en edad preescolar no están habituados a trabajar con limitaciones de tiempo.

Esta prueba consta de 17 ítems, 13 de ellos proceden de la WPPSI-III y 4 son nuevos, incluido el ítem con forma de rombo, creado para ampliar el techo de la prueba. El ítem de ejemplo de la parte A procede de la WPPSI-III, pero se ha añadido un nuevo ítem de ejemplo, en la parte B, para que el niño se familiarice con los cubos de dos colores.

Rompecabezas

Es una prueba de la escala Visoespacial para las edades de 2:6 a 7:7 años. El niño debe unir las piezas de un rompecabezas para reproducir un objeto que se le nombra, dentro de un tiempo límite. El objetivo de esta prueba es medir la organización visoperceptiva, la integración y síntesis de las relaciones parte-todo, el razonamiento no verbal y el aprendizaje por ensayo-error. También se relaciona con la aptitud espacial, la coordinación visomotora, la flexibilidad cognitiva y la persistencia (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008b).

Esta prueba consta de 13 ítems, 11 de ellos proceden de la WPPSI-III y 2 son nuevos. Estos se crearon para mejorar el suelo y el techo de la prueba.

Pruebas de Razonamiento fluido

Matrices

Es una prueba de Razonamiento fluido para las edades de 4:0 a 7:7 años. El niño observa una matriz incompleta y debe seleccionar la opción de respuesta que completa la matriz. El objetivo de esta prueba es medir la inteligencia fluida, la inteligencia visual general, la aptitud espacial y la clasificación, el conocimiento de las relaciones parte-todo, el procesamiento simultáneo y la organización perceptiva (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008b).

De los cuatro tipos de ítems que formaban esta prueba en la WPPSI-III, solo se conservó uno, el de razonamiento analógico. Esta reducción permite un aprendizaje más eficaz de la tarea. Se eliminaron los ítems de patrón continuo y discreto, de clasificación y de razonamiento serial, de los cuales había pocos.

Esta prueba consta de 26 ítems, 21 de ellos proceden de la WPPSI-III y 5 son nuevos. Los 3 ítems de ejemplo también proceden de la WPPSI-III.

Conceptos

Es una prueba de Razonamiento fluido para las edades de 4:0 a 7:7 años. El niño observa dos o tres filas de dibujos y debe elegir un dibujo de cada fila para formar un grupo con una característica común. El objetivo de esta prueba es medir el razonamiento fluido e inductivo, el reconocimiento y procesamiento visoperceptivos y el pensamiento conceptual (Lichtenberger y Kaufman, 2003; Sattler, 2008b); así como con el conocimiento cristalizado (Sattler, 2008b).

Esta prueba consta de 27 ítems, 22 de ellos proceden de la WPPSI-III (algunos de estos ítems se revisaron considerablemente) y 5 son nuevos. Los 2 ítems de ejemplo también proceden de la WPPSI-III. Algunos dibujos se dibujaron de nuevo, y no hay ningún dibujo que aparezca más de una vez a lo largo de la prueba.

Pruebas de Memoria de trabajo

Reconocimiento

Es una prueba de Memoria de trabajo para las edades de 2:6 a 7:7 años. El niño observa una página de estímulos con uno o más dibujos durante un tiempo límite. Luego debe seleccionar los dibujos que ha visto de entre las opciones de una página de respuesta. Esta prueba se asemeja a otras tareas de memoria de trabajo visual (p. ej., Hartshorne, 2008; Makovski y Jiang, 2008), pero es relativamente innovadora porque los estímulos tienen un significado para los niños pequeños. Al igual que en otras tareas de memoria de trabajo, los requisitos de esta prueba se basan en la interferencia proactiva en vez de en la secuenciación.

Reconocimiento es una prueba nueva, creada para la WPPSI-IV, y consta de 35 ítems.

Localización

Es una prueba de Memoria de trabajo para las edades de 2:6 a 7:7 años. El niño observa una o más tarjetas de animales colocadas sobre un plano durante un tiempo límite. Después debe colocar cada tarjeta en el lugar en que estaba. Esta prueba se asemeja a otras tareas de memoria de trabajo visual (Blalock y McCabe, 2011; Lipinski, Simmering, Johnson y Spencer, 2010), pero es relativamente innovadora porque, como en *Reconocimiento*, el contexto y los estímulos de la prueba tienen un significado para los niños pequeños. Para aumentar la carga de la memoria de trabajo a lo largo de los ítems, esta prueba usa la interferencia proactiva, al igual que en *Reconocimiento*.

Localización es una prueba nueva, creada para la WPPSI-IV, y consta de 20 ítems.

Pruebas de Velocidad de procesamiento

Búsqueda de animales

Es una prueba de Velocidad de procesamiento para las edades de 4:0 a 7:7 años. El niño observa un estímulo objetivo (un animal) y un grupo de búsqueda (cinco animales). Después, en un tiempo límite, el niño debe marcar el animal en el grupo de búsqueda que coincide con el estímulo objetivo. Esta prueba es conceptualmente similar a *Búsqueda de símbolos* de la WPPSI-III porque mide la velocidad perceptiva, la memoria visual a corto plazo, la coordinación visomotora, la flexibilidad cognitiva, la discriminación visual y la concentración (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008b).

Los estímulos de *Búsqueda de animales* son imágenes sencillas y grandes, sugerentes para los niños pequeños. Al indicar las respuestas mediante un tampón en vez de un lápiz, los requisitos de la motricidad fina son menores. Se mejoró el diseño distribuyendo más sistemáticamente los estímulos objetivo y los distractores dentro de las series de ítems. Además, se mejoró el suelo de la prueba para que la tarea sea más adecuada al desarrollo de los niños pequeños. Por ejemplo, los niños pequeños que no han alcanzado los niveles de representación simbólica suficientes a menudo no entienden, aunque se les explique y demuestre, que han de marcar la palabra *no* para indicar que el estímulo objetivo no figura en el grupo de búsqueda (es decir, no hay coincidencia). Esta situación de no coincidencia se eliminó; por tanto, todos los animales objetivo coinciden con un animal de los grupos de búsqueda. Cada fila figura de distinto color para indicar más claramente que cada grupo de búsqueda está formado por una única fila. Además, es el examinador quien pasa las páginas, en vez del niño, a fin de reducir aún más los requisitos motores.

Búsqueda de animales es una prueba nueva, creada para la WPPSI-IV, y consta de 66 ítems.

Cancelación

Es una prueba de Velocidad de procesamiento para las edades de 4:0 a 7:7 años. El niño debe buscar y marcar, en un tiempo límite, unos objetos determinados de entre un conjunto de objetos y animales distribuido de forma estructurada o de forma aleatoria. El objetivo de esta prueba es medir la velocidad perceptiva, la velocidad en realizar la prueba, la velocidad de procesamiento visual y funcionamiento mental, el rastreo visual y el reconocimiento y la discriminación visoperceptiva (Flanagan y Kaufman, 2009; Sattler, 2008b). También se relaciona con la atención, la concentración y el recuerdo visual (Sattler, 2008a). Las pruebas de cancelación se han utilizado muy frecuentemente en el campo neuropsicológico como medidas de negligencia visual, inhibición de respuesta y perseveración motora (Lezak, Howieson y Loring, 2004).

Cancelación es una adaptación de la prueba *Animales* de la WISC-IV; en esta nueva versión los estímulos son relevantes y sugerentes para los niños, por lo que resulta más adecuada a su edad. Al indicar las respuestas mediante un tampón en vez de un lápiz, los requisitos de la motricidad fina son menores. Se mejoró el diseño distribuyendo más sistemáticamente los estímulos objetivo que ha de buscar y los distractores en los cuadrantes de la página, asignando a los distractores colores similares a los de los objetivos y aumentando el espacio entre todos los elementos para facilitar que el niño marque los objetivos sin marcar, involuntariamente, los distractores.

Cancelación es una prueba nueva, creada para la WPPSI-IV, y consta de 2 ítems.

Clave de figuras

Es una prueba de Velocidad de procesamiento para las edades de 4:0 a 7:7 años. Utilizando una clave y en un tiempo límite, el niño debe marcar las figuras que se corresponden con el dibujo de cada animal. Esta prueba es una tarea de pares asociados con limitación de tiempo, conceptualmente similar a la prueba *Claves* de la WPPSI-III porque mide la velocidad perceptiva, la memoria a corto plazo, el aprendizaje y el rastreo visual, la flexibilidad cognitiva, la atención y la motivación (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008b). Al indicar las respuestas mediante un tampón en vez de tener que dibujar los símbolos con un lápiz, los requisitos de la motricidad fina son menores. En esta prueba, al igual que en *Búsqueda de animales*, es el examinador quien pasa las páginas, en vez del niño,

Clave de figuras es una prueba nueva, creada para la WPPSI-IV, y consta de 72 ítems.

Terminología de las pruebas

En función de su uso en el cálculo de las puntuaciones compuestas, las pruebas de la WPPSI-IV pueden agruparse en tres categorías: principales, complementarias y opcionales. Para crear las puntuaciones compuestas y la información normativa que figura en las tablas de los anexos de ambos manuales, se utilizaron las pruebas principales, a menos que se especifique de otra forma. Las pruebas complementarias pueden aplicarse junto con las pruebas principales para obtener una representación más amplia del funcionamiento intelectual, así como información adicional para la toma de decisiones clínicas. En algunas ocasiones, cuando no se dispone del resultado de una prueba principal o la prueba se considera inválida, las pruebas complementarias pueden utilizarse como sustitutas de las pruebas principales para calcular las puntuaciones compuestas. Las pruebas opcionales, del mismo modo que las pruebas complementarias, ofrecen una evaluación más amplia del funcionamiento intelectual; sin embargo, a diferencia de las pruebas complementarias, no pueden utilizarse como sustitutas de las pruebas principales para calcular las puntuaciones compuestas. En ocasiones, la clasificación de las pruebas en principales, complementarias y opcionales varía en función de la franja de edad y de la puntuación compuesta. Por ejemplo, *Dibujos* es una prueba principal cuando se utiliza para calcular el Índice de comprensión verbal en niños de 2:6 a 3:11 años; sin embargo, para los niños de 4:0 a 7:7 años, es una prueba opcional que no puede utilizarse para calcular el Índice de comprensión verbal. Debido a que la sustitución de pruebas principales por pruebas complementarias puede introducir un error de medida añadido, existen limitaciones en este procedimiento que se describen en detalle en el apartado «Sustitución de pruebas principales por complementarias» del capítulo 2 del *Manual de aplicación y corrección.*

Contenido y denominación de las puntuaciones compuestas y las puntuaciones de procesamiento

Puntuaciones compuestas

Las puntuaciones compuestas de la WPPSI-IV y sus abreviaciones se presentan en la tabla 2.3. Las diez puntuaciones compuestas permiten describir el rendimiento del niño en dominios del funcionamiento cognitivo más amplios (índices y CI total) que los medidos por las pruebas. Debido a las diferencias en las pruebas incluidas en cada franja de edad, para los niños de 2:6 a 3:11 años se dispone de siete puntuaciones compuestas, mientras que para los niños de 4:0 a 7:7 años se dispone de diez puntuaciones compuestas (véase el apartado «Estructura de la WPPSI-IV» para consultar los detalles sobre las puntuaciones compuestas disponibles para cada franja de edad).

Tabla 2.3. Abreviaciones de las puntuaciones compuestas de la WPPSI-IV

Puntuación compuesta	Abreviación
Índice de comprensión verbal	ICV
Índice visoespacial	IVE
Índice de razonamiento fluido	IRF
Índice de memoria de trabajo	IMT
Índice de velocidad de procesamiento	IVP
CI total	CIT
Índice de adquisición de vocabulario	IAV
Índice no verbal	INV
Índice de capacidad general	ICG
Índice de competencia cognitiva	ICC

Terminología de las puntuaciones compuestas

La WPPSI-IV incorpora cambios importantes en la denominación de las puntuaciones compuestas. Estos cambios reflejan de forma más precisa la estructura de las pruebas y las aptitudes cognitivas evaluadas por cada puntuación compuesta. Esta actualización también conlleva que la WPPSI-IV sea más consistente con las recientes revisiones de las escalas de inteligencia de Wechsler (*Escala de inteligencia de Wechsler para adultos-IV* [WAIS-IV; Wechsler, 2008, 2012] y la WISC-IV). Los antiguos CI verbal (CIV) y CI manipulativo (CIM) se sustituyeron por los términos *Índice de comprensión verbal* (ICV) e *Índice visoespacial* (IVE), respectivamente. Al igual que el CIV, el ICV incluye pruebas que miden aptitudes verbales de razonamiento, comprensión y expresión. En cuanto al *Índice visoespacial*, se optó por utilizar este término porque representaba mejor las aptitudes medidas por las pruebas que lo componen (*Cubos* y *Rompecabezas*). Además, el término *visoespacial* enfatiza y diferencia qué mide esta puntuación compuesta en comparación con el nuevo Índice de razonamiento fluido (IRF) para las edades comprendidas entre los 4:0 y 7:7 años. El término *Velocidad de procesamiento* se sustituyó por *Índice de velocidad de procesamiento* para mantener la consistencia entre la WPPSI-IV y las escalas de inteligencia de Wechsler más recientes. El término *Lenguaje general* de la WPPSI-III pasó a denominarse *Índice de adquisición de vocabulario* para representar de forma más adecuada el constructo que evalúa. La nueva denominación de las puntuaciones compuestas pretende ser más descriptiva y precisa para que el profesional pueda comunicar de una forma clara los resultados del rendimiento del niño a los padres, tutores, profesores u otros profesionales.

Los nueve índices disponibles en la WPPSI-IV pueden dividirse en dos categorías: primarios y secundarios. Los índices primarios son puntuaciones compuestas basadas en los factores que habitualmente se obtienen en una evaluación completa de la aptitud cognitiva, y en esta escala son cinco: el Índice de comprensión verbal, el Índice visoespacial, el Índice de razonamiento fluido, el Índice de memoria de trabajo y el Índice de velocidad de procesamiento. Los índices secundarios tienen principalmente una base teórica y son los cuatro siguientes: el Índice de adquisición de vocabulario, el Índice no verbal, el Índice de capacidad general y el Índice de competencia cognitiva. Estos índices secundarios permiten obtener información adicional o evidencias en relación con el rendimiento de un niño en la WPPSI-IV (véase el capítulo 7 y los anexos B y C de este manual para ampliar la información acerca de la interpretación de los índices secundarios).

Puntuaciones de procesamiento

El enfoque de procesamiento de Boston, desarrollado por Edith Kaplan (1988), destaca que la evaluación cualitativa del rendimiento en un test (análisis de los errores, registro de las observaciones conductuales y evaluación de los límites del sujeto) es tan importante como la evaluación cuantitativa de las puntuaciones. Las revisiones más recientes de las escalas de inteligencia de Wechsler, como la WISC-IV y la WAIS-IV incluyen puntuaciones de procesamiento para proporcionar información adicional acerca del rendimiento del sujeto en una prueba. Para las edades comprendidas entre los 4:0 y 7:7 años se dispone de dos puntuaciones de procesamiento, correspondientes a la prueba *Cancelación: Cancelación aleatoria* (CAA) y *Cancelación estructurada* (CAE). Estas dos puntuaciones de procesamiento se obtienen, respectivamente, de los dos ítems de *Cancelación* sin necesidad de realizar una nueva aplicación. Un resultado poco habitual y significativamente inferior en la puntuación CAA respecto a la puntuación CAE puede indicar que, ante tareas similares, el hecho de disponer de una estructura organizada favorece el rendimiento del niño. En la WPPSI-IV no se dispone de puntuaciones de procesamiento para las edades de 2:6 a 3:11 años.

Estructura de la WPPSI-IV

Debido a los cambios evolutivos que se producen en las aptitudes cognitivas entre los 2:6 y los 7:7 años, el ámbito de aplicación de la WPPSI-IV se divide en dos franjas de edad, de 2:6 a 3:11 años y de 4:0 a 7:7 años, con sendas baterías de pruebas. Las figuras 2.1 y 2.2 representan la estructura de la WPPSI-IV para cada franja de edad. (En el capítulo 6 del este manual puede consultarse la validez de la estructura de la escala a partir de los resultados del análisis factorial.)

En ambas franjas de edad, la estructura de la WPPSI-IV se organiza en tres niveles de interpretación: escala total, escalas primarias y escalas secundarias. Cada nivel de interpretación se compone, a su vez, de escalas (p. ej., escala Comprensión verbal, escala No verbal) formadas por pruebas principales, que se utilizan para obtener la información normativa de la puntuación compuesta, y pruebas complementarias, que pueden sustituir a las pruebas principales que no se han podido aplicar por cualquier motivo o que se han considerado inválidas. En dichas figuras, las pruebas principales de cada escala aparecen en rojo, y las pruebas complementarias, en negro y cursiva. Las pruebas que forman parte de cada escala se utilizan para calcular los índices.

Generalmente, los índices que se obtienen de las escalas primarias (ICV, IVE, IRF, IMT e IVP) se utilizan para evaluar y describir de forma comprehensiva las aptitudes cognitivas del niño. Los índices que se obtienen de las escalas secundarias (IAV, INV, ICG e ICC) se utilizan cuando es necesario complementar los índices primarios en situaciones clínicas (p. ej., la evaluación de niños con dificultades o retraso significativo del lenguaje). Los índices primarios se basan en los resultados del análisis factorial, mientras que los índices secundarios provienen de grupos de pruebas teóricamente relacionados. La información normativa y de interpretación de los índices primarios se encuentra en los anexos A y B del *Manual de aplicación y corrección*. La información correspondiente a los índices secundarios se encuentra en los anexos B y C de este manual.



Figura 2.1. Estructura de la WPPSI-IV para las edades de 2:6 a 3:11 años

Edades de 2:6 a 3:11 años

La WPPSI-IV para las edades de 2:6 a 3:11 años se estructura en tres escalas primarias: Comprensión verbal, Visoespacial y Memoria de trabajo. Las pruebas que forman parte de cada escala primaria se utilizan para calcular el índice primario correspondiente: Índice de comprensión verbal, Índice visoespacial, Índice de memoria de trabajo. La Escala total incluye las pruebas que componen las escalas primarias, así como cualquier prueba complementaria que pueda utilizarse para obtener el CI total (*Nombres*).

Seis pruebas principales componen las escalas primarias, dos pruebas principales para cada escala: *Dibujos* e *Información* forman el ICV; *Cubos* y *Rompecabezas*, el IVE; y *Reconocimiento* y *Localización*, el IMT. No puede utilizarse ninguna prueba complementaria para obtener los índices primarios.

La Escala total está compuesta por cinco pruebas principales: *Dibujos, Información, Cubos, Rompecabezas* y *Reconocimiento*; y dos pruebas complementarias: *Localización* y *Nombres*. Nombres puede sustituir a *Dibujos* pero no a *Información*; y *Localización* puede sustituir a *Reconocimiento*. Es importante tener en cuenta que solo se permite una única sustitución de una prueba principal por

una prueba complementaria para obtener el CI total. En el capítulo 2 del *Manual de aplicación y corrección* se encuentran las normas de sustitución para la obtención de las puntuaciones compuestas.

Para las edades de 2:6 a 3:11 años, la WPPSI-IV cuenta con tres escalas secundarias: Adquisición de vocabulario, No verbal y Capacidad general. Las pruebas que forman parte de cada escala secundaria se utilizan para calcular el índice secundario correspondiente: Índice de adquisición de vocabulario, Índice no verbal e Índice de capacidad general.

Las escalas secundarias están compuestas por un número distinto de pruebas. El IAV está formado por dos pruebas principales: *Dibujos* y *Nombres*, y el INV está formado por cuatro pruebas principales: *Cubos*, *Rompecabezas*, *Reconocimiento* y *Localización*. En esta franja de edad no puede utilizarse ninguna prueba complementaria para obtener el IAV ni el INV. Por último, el ICG está formado por cuatro pruebas principales: *Dibujos*, *Información*, *Cubos* y *Rompecabezas*; y una prueba complementaria: *Nombres*, que puede sustituir a *Dibujos* pero no a *Información*.



Figura 2.2. Estructura de la WPPSI-IV para las edades de 4:0 a 7:7 años

Edades de 4:0 a 7:7 años

La WPPSI-IV para las edades de 4:0 a 7:7 años se estructura en cinco escalas primarias: Comprensión verbal, Visoespacial, Razonamiento fluido, Memoria de trabajo y Velocidad de procesamiento. Las pruebas que forman parte de cada escala primaria se utilizan para calcular el índice primario correspondiente: Índice de comprensión verbal, Índice visoespacial, Índice de Razonamiento fluido, Índice de memoria de trabajo e Índice de velocidad de procesamiento. La Escala total incluye las pruebas que componen las escalas primarias, así como cualquier prueba complementaria que pueda utilizarse para obtener el CI total.

Diez pruebas principales componen las escalas primarias, dos pruebas principales para cada escala: *Información* y *Semejanzas* forman el ICV; *Cubos* y *Rompecabezas*, el IVE; *Matrices* y *Conceptos*, el IRF; *Reconocimiento* y *Localización*, el IMT; y *Búsqueda de animales* y *Cancelación*, el IVP. Al igual que para las edades de 2:6 a 3:11 años, no puede utilizarse ninguna prueba complementaria para obtener los índices primarios.

La Escala total está compuesta por seis pruebas principales: *Información, Semejanzas, Cubos, Matrices, Reconocimiento* y *Búsqueda de animales*; y siete pruebas complementarias: *Vocabulario, Comprensión, Rompecabezas, Conceptos, Localización, Cancelación* y *Clave de figuras*. Las pruebas *Vocabulario* y

Comprensión pueden sustituir tanto a Información como a *Semejanzas*, pero no a ambas; *Rompecabezas* puede sustituir a *Cubos*; *Conceptos* puede sustituir a *Matrices*; *Localización* puede sustituir a *Reconocimiento*; *Cancelación* o *Clave de figuras* pueden sustituir a *Búsqueda de animales*.

Para las edades de 4:0 a 7:7 años, la WPPSI-IV cuenta con cuatro escalas secundarias: Adquisición de vocabulario, No verbal, Capacidad general y Competencia cognitiva. Las pruebas que forman parte de cada escala secundaria se utilizan para calcular el índice secundario correspondiente: Índice de adquisición de vocabulario, Índice no verbal, Índice de capacidad general e Índice de competencia cognitiva.

Las escalas secundarias están compuestas por un número distinto de pruebas. El IAV está formado por dos pruebas principales: *Dibujos* y *Nombres*; y, al igual que para las edades de 2:6 a 3:11 años, no puede utilizarse ninguna prueba complementaria para obtener este índice. El INV está formado por cinco pruebas principales: *Cubos, Matrices, Conceptos, Reconocimiento* y *Búsqueda de animales*; y cuatro pruebas complementarias: *Rompecabezas, Localización, Cancelación* y *Clave de figuras*, que pueden utilizarse como sustitutas cuando no se ha podido aplicar una prueba principal o la prueba se considera inválida. El ICG está formado por cuatro pruebas principales: *Información, Semejanzas, Cubos* y *Matrices*; y cuatro pruebas complementarias: *Vocabulario, Comprensión, Rompecabezas* y *Conceptos*. El ICC está formado por cuatro pruebas principales: *Reconocimiento, Localización, Búsqueda de animales* y *Cancelación*; y una prueba complementaria: *Clave de figuras*, que, cuando es necesario, puede sustituir a *Búsqueda de animales* o a *Cancelación*.

Es importante tener en cuenta que solo se permite una única sustitución de una prueba principal por una prueba complementaria para obtener el CI total y los índices secundarios. Además, solo se permiten sustituciones entre pruebas de un mismo dominio cognitivo primario (Comprensión verbal, Visoespacial, Razonamiento fluido, Memoria de trabajo y Velocidad de procesamiento). Por ejemplo, si es necesario, se puede utilizar la prueba *Vocabulario* para sustituir a otra prueba de Comprensión verbal, pero no se puede utilizar *Vocabulario* para sustituir a una prueba inválida de Razonamiento fluido, Memoria de trabajo, Visoespacial o Velocidad de procesamiento. En el capítulo 2 del *Manual de aplicación y corrección* se encuentran las normas de sustitución para la obtención de las puntuaciones compuestas.

Objetivos de la revisión

Los objetivos de esta revisión se establecieron tras considerar distintos modelos estructurales de la inteligencia y de la capacidad cognitiva contemporáneos, la investigación sobre el neurodesarrollo y la neurocognición, los resultados psicométricos, la utilidad clínica y las necesidades prácticas de los profesionales y de la sociedad. Asimismo se basan en setenta años de investigación con las escalas de inteligencia de Wechsler, y se establecieron a partir de las aportaciones de expertos en los campos de la neuropsicología clínica infantil, del desarrollo cognitivo y del lenguaje de los niños; de las sugerencias de los psicólogos que usan las escalas y, también, del amplio estudio de la literatura existente en los campos de la teoría e investigación de la capacidad cognitiva, la neurociencia cognitiva, la evaluación intelectual y el desarrollo cognitivo de los niños. Los cinco objetivos principales fueron los siguientes:

- actualización de las bases teóricas;
- mayor adecuación al desarrollo;
- aumento de la facilidad de uso;
- mejorar las propiedades psicométricas; e
- incremento de la utilidad clínica.

Actualización de las bases teóricas

En este apartado se exponen los modelos estructurales de la inteligencia contemporáneos, las investigaciones sobre el neurodesarrollo y la neurocognición, los modelos sobre la memoria de trabajo y los estudios empíricos relacionados con el desarrollo de la WPPSI-IV.

Modelos estructurales de la inteligencia contemporáneos

Las ideas de Wechsler sobre la inteligencia y las progresivas adaptaciones de sus escalas de inteligencia a las teorías e investigaciones contemporáneas parecen contrarias a la afirmación de que sus escalas carecen de base teórica (Coalson, Raiford, Saklofske y Weiss, 2010; Kaufman, 2010). No obstante, algunos autores afirman o insinúan que las escalas de inteligencia de Wechsler carecen de estructura o contexto teórico (Benson, Hulac y Kranzler, 2010; Kamphaus, 1993; Keith, Fine, Taub, Reynolds y Kranzler, 2006). Cuando Wechsler creó la Wechsler-Bellevue, destacaban dos teorías de la inteligencia: la de Spearman (1904), quien defendía la existencia de un factor general y otros específicos, y la de Thorndike (1936), que consideraba que hay varios tipos de inteligencia, como la abstracta, la social y la práctica. Aunque se consideró a Wechsler como un teórico del factor *g* (Kaufman, Flanagan, Alfonso y Mascolo, 2006), quizá por incluir una puntuación resumida para representar la inteligencia general, su definición de inteligencia (véase el capítulo 1 de este manual) no es incompatible ni con la teoría del factor *g* ni con la de los distintos tipos de inteligencia. Wechsler describió un aspecto general de la inteligencia, compuesto por aptitudes distintas cualitativamente (Coalson et al., 2010), y, cuando desarrolló sus tests, seleccionó medidas de la capacidad cognitiva que luego resultaron ser importantes según los modelos estructurales de la inteligencia contemporáneos (Carroll, 1993, 2012). Además, incorporó el uso de puntuaciones compuestas alternativas basadas en los estudios del análisis factorial (véanse Cohen, 1957a, 1957b, 1959 y Kaufman, 1975) y en los resultados de estudios clínicos que indicaban su utilidad en el diagnóstico diferencial (Coalson et al., 2010; Wechsler, 1958).

Actualmente, los modelos estructurales de la inteligencia ampliamente aceptados se basan en resultados exhaustivos del análisis factorial. Estos estudios demuestran con firmeza que hay una inteligencia general en el nivel superior de un modelo jerárquico, y varias aptitudes generales relacionadas y distinguibles en el nivel inferior. En algunos modelos, las aptitudes específicas están compuestas de varias aptitudes más segmentadas en el nivel más inferior (Carroll, 1993, 2012; Horn y Blankson, 2012; Johnson et al., 2004; Salthouse, 2004). Aunque las evidencias aportadas por los modelos estructurales no coinciden completamente, la mayoría indican que las aptitudes de comprensión verbal, visoespacial, razonamiento fluido, memoria de trabajo y velocidad de procesamiento figuran entre los componentes fundamentales (Carroll, 1993, 2012; Horn y Blankson, 2012; Johnson, te Nijenhuis y Bouchard 2007; Salthouse, 2004). Las escalas de inteligencia de Wechsler (Wechsler, 1991, 1997, 2002, 2003, 2008) han evolucionado consecuentemente, incorporando nuevas medidas de razonamiento fluido, memoria de trabajo y velocidad de procesamiento. La WPPSI-IV continúa esta línea evolutiva ofreciendo un nuevo índice de memoria de trabajo, proporcionando un índice de razonamiento fluido y un índice visoespacial independientes, y mejorando las medidas de comprensión verbal y velocidad de procesamiento, manteniendo las puntuaciones compuestas para ambas.

La teoría no es el único elemento que guía el desarrollo de las escalas de inteligencia de Wechsler, pero el modelo de la WPPSI-IV refleja las teorías estructurales contemporáneas, y la interpretación de la escala puede apoyarse en estas teorías (p. ej., Flanagan y Kaufman, 2009). Queda fuera del alcance de este capítulo exponer con detalle las teorías de la inteligencia contemporáneas; si se desea más información sobre los modelos que pueden aplicarse al interpretar la WPPSI-IV, véanse Flanagan y Harrison (2012) y Sattler (2008a).

Investigación sobre el neurodesarrollo y la neurocognición

En la evaluación neuropsicológica clínica se utilizan tests cognitivos con el objetivo de determinar algún deterioro o cambio posterior a un traumatismo o a una enfermedad, en especial relacionados con determinadas estructurales cerebrales o vías neurales. Varios de los constructos cognitivos fundamentales medidos por la WPPSI-IV pueden utilizarse para generar hipótesis sobre déficits del procesamiento neuropsicológico, y pueden interpretarse desde la perspectiva del neurodesarrollo (Maricle y Avirett, 2012; McCloskey, Whitaker, Murphy y Rogers, 2012; Miller y Maricle, 2012). Debido a aspectos metodológicos, pocos estudios de neuroimagen examinan la activación de las regiones cerebrales durante la realización de tests de inteligencia; por lo tanto, no se ha relacionado explícitamente la activación regional del cerebro con la aplicación de las pruebas de estos tests. Este tipo de investigación puede llevarse a cabo más fácilmente con adultos que con niños. Un estudio reciente empleó técnicas de cartografía de las lesiones (*lesion-mapping*)

para evaluar la activación cerebral durante la aplicación de la WAIS-III. Los resultados obtenidos indicaron que se activaron regiones cerebrales más localizadas para el ICV, el IOP y el IMT que para el IVP, que parece ser más difuso (Gläscher et al., 2009). Otros estudios relacionados vinculan tanto la inteligencia general como las capacidades cognitivas específicas medidas por la WPPSI-IV con aspectos estructurales y funcionales del desarrollo cerebral. En atención a los avances realizados en el campo de las imágenes cerebrales, los psicólogos y los educadores deberían revisar sus conocimientos actuales sobre la inteligencia a fin de incluir la literatura relevante acerca de las bases neurales de la inteligencia (Haier y Jung, 2008).

Neurodesarrollo e inteligencia general

El cerebro se desarrolla rápidamente durante la primera infancia. La materia blanca, que incluye la mielina (una sustancia que aumenta la velocidad de la transmisión de los impulsos nerviosos por los axones [Hale y Fiorello, 2004]), se incrementa gradualmente (Lebel et al., 2012; Tamnes et al., 2011); y la materia gris (neuronas) se incrementa de forma no lineal (Gogtay y Thompson, 2010; Pangelinan et al., 2011; Shaw et al., 2006). El desarrollo del cerebro en la primera infancia se caracteriza por la proliferación de sinapsis seguida de poda sináptica. La poda sináptica es un proceso mediante el cual se crean conexiones sinápticas entre neuronas al mismo tiempo que se destruyen otras, con el fin de reforzar el funcionamiento cerebral óptimo y de eliminar los circuitos neuronales ineficientes (Hale y Fiorello, 2004). La mielinización y el fortalecimiento de las conexiones entre las regiones cerebrales se produce a lo largo de la primera infancia a una velocidad extraordinaria (de Bie et al., 2012; Lee, Wu, Yu, Wu y Chen, 2012). En conjunto, estos cambios representan el perfeccionamiento de las estructuras neurales y de las vías de fibras responsables de los procesos cognitivos de nivel superior; por ello, no debe sorprender que numerosos estudios hayan demostrado que la inteligencia general está relacionada con el desarrollo de la materia blanca y la materia gris cerebrales (Dumontheil, Houlton, Christoff y Blakemore, 2010; Pangelinan et al., 2011; Yuan et al., 2012). Además, estos estudios indican que la inteligencia general se relaciona con una serie de regiones neurales y las vías que las conectan, lo cual es consistente con la naturaleza multifacética de la aptitud cognitiva.

Neurodesarrollo y aptitudes cognitivas específicas

Las puntuaciones de los índices primarios de la WPPSI-IV representan las aptitudes cognitivas específicas que han sido objeto de exhaustivas investigaciones científicas a fin de descubrir sus asociaciones con el desarrollo cerebral. Tales investigaciones son de gran interés para los profesionales puesto que proporcionan información sobre el desarrollo cerebral normal y el desarrollo cerebral anormal asociado a distintos trastornos clínicos (Hale y Fiorello, 2004). Esta información, además, puede ser de ayuda para establecer el diagnóstico y planificar el tratamiento.

La primera infancia se caracteriza por el enorme desarrollo de las habilidades verbales. El desarrollo del lenguaje, que incluye las aptitudes de comprensión auditiva, varía considerablemente de un niño a otro. La fonología (sonidos), la semántica (significado de las palabras y vocabulario) y la sintaxis (gramática) interactúan dinámicamente e influyen en el desarrollo del lenguaje (Kuhl et al., 2008). Varias de las regiones cerebrales tradicionalmente asociadas al habla muestran su activación durante las tareas en las que interviene la fonología, la semántica y/o la sintaxis (Bornkessel-Schlesewsky, Grewe, y Schlesewsky, 2012; Christensen, 2009; Price, 2000). Aunque las pruebas de Comprensión verbal de la WPPSI-IV están relacionadas con tareas lingüísticas, también guardan una estrecha relación con la inteligencia general.

La aptitud visoespacial mejora durante la primera infancia conforme maduran la materia blanca y la materia gris de la zona parietal (Olesen, Nagy, Westerberg y Klingberg, 2003). Estudios realizados mediante imagen funcional han vinculado la cognición visoespacial con la activación de los lóbulos parietales tanto en niños como en adultos (Ganis, Thompson y Kosslyn, 2004; Heil y Jansen-Osmann, 2007; Kosslyn, Thompson y Alpert, 1997). Asimismo otras regiones cerebrales pueden estar asociadas con determinadas tareas espaciales o visuales. Por ejemplo, en la región occipital media izquierda se ha asociado la densificación de la materia gris con la obtención de mejores resultados por parte de los niños pequeños en la prueba *Cubos* (Sowell et al., 2004).

El razonamiento fluido, un componente importante de la aptitud intelectual, se asocia con la activación del córtex prefrontal lateral en los niños pequeños (Kuwajima y Sawaguchi, 2010). La

activación del córtex prefrontal lateral también se asocia con la función ejecutiva y la memoria de trabajo (Kuwajima y Sawaguchi, 2010), ambas estrechamente relacionadas con el razonamiento fluido (Martínez et al., 2011; Schneider et al., 2005).

La memoria de trabajo se asocia con el córtex prefrontal dorsolateral y ventrolateral (Jolles, Kleibeuker, Rombouts y Crone, 2011) y otras áreas del córtex prefrontal que maduran rápida y constantemente durante la infancia. Otros aspectos del desarrollo cerebral se relacionan con el rendimiento en tareas de memoria de trabajo visual. En general, los estudios sobre el neurodesarrollo de la memoria de trabajo sugieren que las diferencias en el rendimiento a lo largo de la primera infancia y la adolescencia reflejan la maduración funcional de numerosas estructuras cerebrales (p. ej., Kwon, Reiss y Menon, 2002; Martin, Houssemand, Schiltz, Burnod y Alexandre, 2008; Nagy, Westerberg y Klingberg, 2004).

Las mejoras en la velocidad de procesamiento, un componente importante de la inteligencia en los niños pequeños (Rose, Feldman y Jankowski, 2009; Rose, Feldman, Jankowski y Van Rossem, 2008), se relacionan con cambios relacionados con la edad, en el número de conexiones temporales del sistema nervioso central e incrementos de la mielinización (Berthier, DeBlois, Poirier, Novak y Clifton, 2000; Cepeda, Kramer y Gonzalez de Sather, 2001). El incremento de la mielinización implica una mejor velocidad de procesamiento, puesto que supone una mayor integración de múltiples regiones cerebrales y una mayor eficiencia del procesamiento neural. Esta estrecha relación de la mielinización y la materia blanca con el rendimiento en tareas de velocidad de procesamiento puede explicar en parte la sólida relación entre la madurez de la materia blanca y la inteligencia general (Barnea-Goraly et al., 2005; Matute y Ransom, 2012; Paus et al., 2001).

Como demuestran estas observaciones, el cerebro de los niños pequeños se ve sometido a un rápido cambio. Dado el desarrollo dinámico de las estructuras cerebrales y de la aptitud cognitiva, caracterizar la aptitud intelectual de los niños pequeños constituye un reto. Los profesionales deben tener en mente que el rendimiento de un niño pequeño cambia de un día a otro y que los rápidos cambios en el neurodesarrollo pueden influir en el rendimiento.

Modelos e investigación sobre la memoria de trabajo

Aunque en la WPPSI-III se eliminaran tres pruebas de memoria porque no tenían un suelo suficientemente bajo, no por ello debe considerarse que la memoria no es un aspecto importante de la inteligencia en los niños pequeños (Wechsler, 2002). La memoria de trabajo es un aspecto clave del funcionamiento intelectual, con fuertes vínculos con el razonamiento fluido (Burgess y Braver, 2010; Hornung, Brunner, Reuter y Martin, 2011); también interviene en una amplia variedad de problemas académicos y trastornos clínicos que afectan a los niños pequeños (p. ej., Borella, Carretti y Pelegrina, 2010; Fitzpatrick y Pagani, 2012; Hutchinson, Bavin, Efron y Sciberras, 2012). Consecuentemente, uno de los objetivos desde el inicio del desarrollo de la WPPSI-IV fue incorporar medidas de memoria de trabajo.

Por lo general, la memoria de trabajo se define como la información que se mantiene en la mente y que es necesaria para llevar a cabo y, al mismo tiempo, completar correctamente algunos tipos de tareas cognitivas (Cowan y Alloway, 2009). Comúnmente, se acepta que la memoria de trabajo existe incluso en los niños más pequeños (de hasta 2 o 3 años), pero medir este tipo de memoria en estas edades tempranas es metodológicamente muy complejo. De hecho, los niños pequeños tienen una capacidad de memoria de trabajo limitada, mucha facilidad para distraerse, dificultad para comprender instrucciones complejas y no están capacitados para usar estrategias de repetición. Aun así, los niños pequeños muestran cierta memoria de trabajo, aunque simplemente consista en retener información frente a una demanda cognitiva concurrente (Reznick, 2009).

Los niños pequeños de 3 a 4 años pueden retener unos dos elementos en la memoria de trabajo; mientras que los niños de 7 años, en promedio, doblan dicha capacidad (Simmering, 2012). Las pruebas de memoria de trabajo de la WPPSI-IV incluyen estos aspectos y están construidas a partir de dos paradigmas descritos por Reznick (2009): *familiarizarse-reconocer y observar-realizar.* En las tareas de familiarizarse-reconocer, los niños observan una serie de estímulos y luego han de reconocerlos de entre otros elementos (como en *Reconocimiento*). En las tareas de observar-realizar, los niños observan una serie de acciones y luego han de reproducirlas (como en *Localización*). Las pruebas de memoria de trabajo de la WPPSI-IV, fundamentadas en estos paradigmas, permiten

minimizar las dificultades que se encuentran en otras tareas; por ejemplo, tareas que requieren instrucciones complejas.

Es deseable que en algunos modelos de memoria de trabajo haya tareas con demandas cognitivas simultáneas y competitivas. Lamentablemente, la complejidad de las instrucciones hace que muchas tareas duales y paradigmas de secuenciación tradicionales que podrían usarse para incrementar las demandas cognitivas concurrentes resulten inadaptables a niños muy pequeños, especialmente cuando el problema que motiva la evaluación neurocognitiva del niño afecta a la comprensión de las instrucciones. Las tareas de memoria de trabajo de la WPPSI-IV, por el contrario, se basan en la interferencia proactiva para introducir una demanda cognitiva concurrente. La interferencia proactiva tiene lugar cuando un ítem visto previamente interfiere con la memoria del ítem presente. Este método fue usado por primera vez por Piaget en sus estudios de la memoria de trabajo con niños de hasta 2 o 3 años. Diversos estudios han demostrado que la incorporación de la interferencia proactiva en un conjunto de ítems de memoria de trabajo incrementa la dificultad y la carga de memoria de trabajo de los ítems posteriores (Lipinski et al., 2010; Makovski y Jiang, 2008; Szmalec, Verbruggen, Vandierendonck y Kemps, 2011).

Aunque queda fuera del alcance de este apartado presentar un análisis completo de los modelos de memoria de trabajo actuales, se ofrece una comparación de dos modelos ampliamente aceptados dado que son relevantes para las pruebas de memoria de trabajo de la WPPSI-IV: por una parte, el modelo multicomponente de la memoria de trabajo y, por otra parte, el modelo de procesos anidados. El modelo multicomponente de la memoria de trabajo establece dos sistemas de almacenamiento específicos según el dominio: el bucle fonológico para la información verbal y la agenda visoespacial para la información visual y espacial (Baddeley, 2000, 2002, 2012). Por otra parte, el modelo de procesos anidados (Cowan, 1988; Towse y Cowan, 2005), que presenta ventajas y respaldo empírico, propone que cada fase de la memoria es un subconjunto de otra fase de la memoria; es decir, que el foco de la atención, activado por los estímulos del entorno, es un subconjunto de la memoria activada (almacén a corto plazo), que a su vez, es un subconjunto del almacen a largo plazo. Estos dos modelos (multicomponente y de procesos anidados) se diferencia en varios aspectos, siendo uno de ellos, y no el menos importante, que el modelo multicomponente propone distinciones en función del dominio (p. ej., verbal y visoespacial), mientras que el modelo de procesos anidados no las establece. Se ha reconocido que incluso las tareas de memoria de trabajo visoespacial pueden implicar algunos procesos de memoria verbal; este hecho demuestra la dificultad en concebir la memoria como exclusivas de un dominio u otro (Towse y Cowan, 2005).

No obstante, los componentes de almacenamiento específicos según el dominio parecen ser distintos en los niños pequeños cuya edad se sitúa en el rango abarcado por la WPPSI-IV. Las investigaciones con niños pequeños respaldan significativamente el modelo multicomponente de Baddeley (Alloway, Gathercole y Pickering, 2006; Hornung et al., 2011). El desarrollo de la memoria de trabajo en estos dominios parece que se produce de forma independiente. En los niños pequeños, el almacenamiento en la memoria dentro del dominio visual está estrechamente vinculado con el procesamiento, mientras que el almacenamiento en la memoria verbal no lo está (Alloway et al., 2006; Hornung et al., 2011). Esto demuestra que, para medir la memoria de trabajo en los niños pequeños, es preferible una tarea simple de memoria visual (que implique solo almacenamiento) a una tarea verbal análoga. Incluso si no se tiene en cuenta la interferencia proactiva implícita en la prueba *Reconocimiento*, el almacenamiento visual simple supone por sí mismo la intervención del procesamiento y, por tanto, mide la memoria de trabajo.

Mayor adecuación al desarrollo

Al evaluar a niños pequeños, hay que tener en cuenta muchos aspectos fundamentales del desarrollo (Brassard y Boehm, 2007; Ford et al., 2012; Kaufman y Kaufman, 1977; Vig y Sanders, 2007). Por ello, en esta nueva versión de la WPPSI se le concedió una importancia primordial a la mejora de la adecuación al desarrollo; en particular, respecto a los materiales, las consignas y las pruebas de velocidad de procesamiento.

Materiales manipulativos

Los materiales manipulativos de la WPPSI-IV (p.ej., tampón, cubos o piezas de puzle) se seleccionaron presentando una especial atención al nivel de desarrollo. Un ejemplo de ello es la elección del tampón como medio para indicar las respuestas en las pruebas de Velocidad de procesamiento; es un utensilio usado en las guarderías, parvularios y escuelas de educación infantil, así como en diversas actividades de ocio para niños pequeños. Se valoraron diversos tipos de tampones, pero se eligió uno de ellos, en parte porque su diámetro es pequeño y, así, a los niños, que tienen las manos pequeñas, les resulta más fácil cogerlo. Los otros elementos manipulativos de la WPPSI-IV (cubos, puzles de colores y tarjetas de juegos de memoria) son también conocidos por los niños pequeños y acordes con su nivel de desarrollo (Kaufman y Kaufman, 1977; Vig y Sanders, 2007).

Otro aspecto que se tuvo en cuenta fue la seguridad de los objetos que manipula el niño en relación con su nivel de desarrollo. A veces los niños utilizan los objetos de una forma no prevista por los adultos. Los componentes de la WPPSI-IV están fabricados conforme a los estándares de seguridad estadounidenses e internacionales. Unos reputados laboratorios externos realizaron completas evaluaciones y pruebas de cada uno de los componentes; proporcionaron las directrices sobre las advertencias de seguridad de los productos; y ratificaron que los productos cumplen todos los estándares aplicables, incluidas las limitaciones para los productos químicos prohibidos o de uso restringido.

Consignas

Un aspecto importante en la evaluación de los niños pequeños es su comprensión de las consignas de las tareas (Brassard y Boehm, 2007; Ford et al., 2012). Todas las consignas de las pruebas se simplificaron en las fases iniciales de desarrollo de la WPPSI-IV. Las consignas se probaron repetidamente con niños de 2:0 a 2:5 años (más pequeños que los niños de la franja de menor de edad de la WPPSI-IV), así como en niños con alguna discapacidad intelectual, a fin de garantizar que los niños con retrasos del desarrollo o con una aptitud intelectual baja podían comprender los requisitos de las tareas. A lo largo del desarrollo de la WPPSI-IV se pidió a los niños que explicaran sus estrategias y sus respuestas, con el objetivo de comprobar que las consignas de las pruebas motivaban los procesos cognitivos previstos. Asimismo se pidió a los examinadores su opinión sobre la claridad de las consignas.

Los ítems de demostración, de ejemplo y de aprendizaje se utilizan en varias pruebas para explicar las tareas. Los ítems de demostración los resuelve el examinador, los ítems de ejemplo los resuelve el niño, y le sirven para entrenarse en la tarea antes de empezar los ítems del test. A menudo también le permiten al examinador, en función del rendimiento del niño, volver a explicar en qué consiste la tarea. Para las pruebas que no cuentan con ítems de demostración ni de ejemplo (p. ej., *Información* y *Vocabulario*), si el niño no ha obtenido puntuaciones perfectas en los ítems de comienzo, que sirven como ítems de aprendizaje, el examinador dispone de estos ítems para explicar al niño cuál es la respuesta correcta.

Las consignas de la WPPSI-IV son breves y concisas y, cuando es posible, se complementan con ítems de demostración y de ejemplo, puesto que los niños pequeños no suelen mantener la atención y concentración cuando se les dan explicaciones largas (Brassard y Boehm, 2007; Ford et al., 2012; Kaufman y Kaufman, 1977). Durante las aplicaciones de la fase piloto se analizó la brevedad de las consignas y se realizaron varias revisiones de las mismas. Si una consigna hacía referencia a una conducta observada en pocas ocasiones o a una pregunta raramente formulada, la consigna se transformaba en una consigna adicional que únicamente se da cuando es necesario. Las consignas de todas las pruebas se leyeron en voz alta y se cronometraron. Además, se contó el número de palabras de cada consigna. Para cada una de las pruebas conservadas de la WPPSI-III, se compararon el tiempo de lectura y el número de palabras de las consignas de la WPPSI-III con los resultados correspondientes de la WPPSI-IV para garantizar que las consignas nuevas requerían menos tiempo y estaban formadas por menos palabras. No obstante, algunas de las consignas verbales se mantuvieron igual porque ya eran adecuadas al desarrollo. Por ejemplo, en *Rompecabezas*, el examinador nombra los objetos representados en los puzles antes de que el niño empiece a encajar las piezas, con el fin de minimizar su frustración (Kaufman y Kaufman, 1977).

Pruebas de Velocidad de procesamiento

Desde las primeras fases de elaboración de la WPPSI-IV, se fijó como objetivo mejorar las pruebas de Velocidad de procesamiento para su mayor adecuación al desarrollo infantil. Se analizaron los procesos de respuesta de las pruebas de la WPPSI-III teniendo en cuenta el nivel de desarrollo de los niños. Este análisis, junto con la revisión de la literatura sobre el desarrollo infantil, mostró que era necesario realizar varios cambios. Así, por ejemplo, las tareas que se asemejan a un juego generan un ambiente favorable para evaluar a los niños pequeños, porque el juego es adecuado a su nivel de desarrollo (Brassard y Boehm, 2007; Kaufman y Kaufman, 1977). Por ello, todas las pruebas de Velocidad de procesamiento de la WPPSI-IV incluyen estímulos sugerentes y conocidos por los niños pequeños, y los requisitos de las tareas evocan juegos infantiles habituales. Además, dado que la comprensión simbólica se encuentra en una fase inicial en los niños pequeños, se priorizaron los estímulos realistas sobre los estímulos abstractos (Piaget, 1952). La literatura sobre el desarrollo también indica que a los niños pequeños, dado que están desarrollando las destrezas de la motricidad fina, les resulta complejo utilizar un lápiz (Brassard y Boehm, 2007). Por ello, en todas las pruebas de Velocidad de procesamiento se utiliza un tampón en vez de un lápiz para indicar las respuestas.

Búsqueda de animales es, probablemente, la prueba con una mayor adecuación al desarrollo en relación con la prueba equivalente de la WPPSI-III, *Búsqueda de símbolos*. Tras realizar un exhaustivo análisis de los procesos de respuesta de esta prueba de la WPPSI-III, se determinó que, incluso con la demostración, los ejemplos y las explicaciones, algunos niños no entendían que debían marcar la palabra *no* para indicar que un símbolo objetivo no se encontraba en el grupo de búsqueda (es decir, situación de no coincidencia). Se realizaron observaciones similares durante las aplicaciones llevadas a cabo con una primera versión de *Búsqueda de animales* que incluía una representación más concreta de la situación de no coincidencia: el niño debía realizar una marca con el tampón sobre una hoja si el animal objetivo no figuraba en el grupo de búsqueda. Ante esta constatación, se determinó que la versión final de *Búsqueda de animales* no incluyera la situación de no coincidencia; por lo tanto, en todos los ítems de esta prueba, todos los animales objetivo siempre tiene su equivalente en el grupo de búsqueda. Dado que algunos niños buscaban el animal objetivo en todos los grupos de búsqueda que figuraban en la página, en vez de en el grupo de búsqueda de la fila correspondiente al ítem, se decidió que cada fila fuera de un color diferente.

Al crear las pruebas *Cancelación* y *Clave de figuras* se consideró el desarrollo cognitivo, motor y lingüístico. Los estímulos de *Cancelación* de la WPPSI-IV son, respecto a los estímulos de la prueba *Animales* de la WISC-IV, más familiares para los niños pequeños y se encuentran más espaciados en la página para reducir la posibilidad de que un niño con problemas motores indique una respuesta incorrecta sin querer. A partir del análisis del nivel léxico de los niños, la consigna de esta prueba hace mención a «cosas que la gente se pone» en vez de a «prendas de vestir», dado que los niños aprenden antes las palabras de la consigna usada que el sustantivo compuesto señalado. *Clave de figuras* también está formada por estímulos y conceptos conocidos para los niños pequeños y presentados con una separación mayor para garantizar que los requisitos de motricidad fina sean mínimos. Los estudios actuales realizados por investigadores neopiagetianos cuantifican el número de ítems que pueden mantenerse activos en la memoria en distintas edades (Pascual-Leone y Johnson, 2011). Por ello, se redujo en consecuencia el número de pares asociados que deben recordarse; *Clave de figuras* incluye tres pares asociados, mientras que la prueba equivalente de la WPPSI-III, denominada *Claves*, incluía cinco.

Aumento de la facilidad de uso

Seguridad de los ítems

La seguridad de los ítems es un aspecto importante para los usuarios de tests de inteligencia porque los ítems que han pasado a ser públicos se practican y se memorizan con más facilidad. Algunos estudios indican que el contenido de un test se considera del dominio público cuando han transcurrido varios años desde su publicación (Demas y Brogdon, 1997). Para garantizar que el contenido de los ítems de la WPPSI-IV no sea público, se cambiaron todos los ítems que figuraban en la WPPSI (1967) o en la WPPSI-R (1989).

Materiales y presentación

Tanto los materiales como su formato de presentación se rediseñaron para facilitar su uso. Se distribuyeron más equitativamente los contenidos de los cuadernos de estímulos para reducir su peso y, así, facilitar su transporte y evitar que la encuadernación se estropease. Hay tres cuadernos de estímulos, pero solo se necesitan dos para obtener todas o la mayoría de las puntuaciones compuestas requeridas para llevar a cabo una evaluación completa. Para los niños de 2:6 a 3:11 años, los cuadernos de estímulos 1 y 2 incluyen todas las pruebas necesarias para obtener todas las puntuaciones compuestas posibles. Para los niños de 4:0 a 7:7 años, los cuadernos de estímulos 2 y 3 incluyen las pruebas necesarias para obtener todas las puntuaciones compuestas posibles excepto el IAV (porque *Dibujos* y *Nombres* solo figuran en el cuaderno de estímulos 1). Se proporcionan dos tampones por si uno de ellos no marca correctamente, se pierde o se estropea. Asimismo, si es necesario, en Pearson Clinical & Talent Assessment pueden adquirirse más tampones.

Tiempo de aplicación

Dado que los niños pequeños tienen dificultad para mantener la atención durante periodos largos, el tiempo de aplicación deber reducirse al mínimo. La reducción del tiempo de aplicación de la WPPSI-IV permite disponer de tiempo para evaluar otros dominios importantes (p. ej., rendimiento académico, lenguaje, función ejecutiva) cuando es necesario. Una práctica adecuada en la evaluación de la primera infancia requiere evaluar varias áreas además de la aptitud cognitiva, como el desarrollo social, motor y lingüístico. La evaluación debería incluir tanto la aplicación de tests estandarizados como la entrevista, la observación y el juego (Brassard y Boehm, 2007; Ford et al., 2012). Por este motivo, los profesionales piden que el tiempo de aplicación sea lo más breve posible, sin dejar de obtener la información necesaria; es decir, piden un instrumento que permita medir adecuadamente diversos dominios cognitivos importantes (Camara, Nathan y Puente, 2000). A lo largo del desarrollo de la WPPSI-IV se realizó un esfuerzo considerable para lograr que el tiempo de aplicación fuera lo más breve posible y que se ofreciera una amplitud del constructo y una flexibilidad mayores, así como más puntuaciones compuestas.

Además de abreviar las consignas de las pruebas, el número de ítems aplicado se redujo al mínimo porque se disminuyó el número total de ítems y se modificaron las reglas de terminación (véase más adelante el apartado «Reglas de terminación», en este mismo capítulo). Para las pruebas procedentes de la WPPSI-III, el número total de ítems se redujo, como promedio, en 2.9 ítems. Aunque la prueba *Comprensión* de la WPPSI-IV cuenta con dos ítems más respecto a la WPPSI-III, en realidad tiene dos ítems verbales menos. Esta diferencia en el número total de ítems se debe a que se añadieron cuatro ítems gráficos, más rápidos de anotar y puntuar que los ítems verbales, y que, por otra parte, permitieron ampliar el suelo de la prueba. Asimismo, *Semejanzas* cuenta con un ítem menos, pero cuatro de los ítems totales son ítems gráficos.

El tiempo de aplicación es un aspecto fundamental al seleccionar las pruebas de los índices primarios y del CI total. Un exhaustivo estudio analizó el efecto de potenciales reducciones en el número de pruebas que conforman las puntuaciones de los índices primarios y del CI total. Para tomar las decisiones finales sobre las pruebas que conforman las puntuaciones compuestas, se utilizaron diversas fuentes; entre ellas, las opiniones de destacados expertos en inteligencia; la evaluación estadística de la utilidad predictiva y clínica de varias combinaciones de pruebas para obtener las puntuaciones compuestas; y las opiniones de profesionales sobre la amplitud del constructo y el tiempo de aplicación considerados adecuados.

Instrucciones de aplicación y corrección

Las instrucciones de aplicación y corrección de la WPPSI-IV son más explícitas y sencillas que las de la WPPSI-III. Por ejemplo, las reglas de invalidez, prorrateo y sustitución de las puntuaciones compuestas se simplificaron considerablemente y son más fáciles de aplicar (véase el capítulo 2 del *Manual de aplicación y corrección* para más información sobre dichas reglas). Otro ejemplo de simplificación son las reglas de repetición. En la WPPSI-IV, la repetición de ítems o de consignas, cuando está permitida, se diferencia claramente de la reexposición de una imagen o de un modelo durante la aplicación.

Reglas de terminación

Para todas las pruebas procedentes de la WPPSI-III se redujeron significativamente las reglas de terminación. Así, por ejemplo, la regla de terminación para Comprensión era, en la WPPSI-III, de 5 respuestas de 0 puntos consecutivas, y en la WPPSI-IV pasó a ser de 3 respuestas de 0 puntos consecutivas. Las reglas de terminación para las pruebas nuevas se establecieron para minimizar el número necesario de respuestas de 0 puntos consecutivas y, al mismo tiempo, garantizar que se aplica un número adecuado de ítems para obtener una estimación precisa del rendimiento. En el capítulo 4 de este manual se ofrece información detallada sobre los estudios empíricos llevados a cabo para determinar las reglas de terminación.

Mejora de las propiedades psicométricas

Ítems y reglas de puntuación

Con el tiempo, los ítems y las reglas de puntuación pueden quedar obsoletos. Por ello se revisaron los ítems y se sustituyeron los anticuados por ítems nuevos que reflejasen mejor las ideas y situaciones actuales. Las imágenes se dibujaron de nuevo para que los niños pequeños pudieran reconocer fácilmente su contenido. Las reglas de puntuación de las pruebas que requieren respuestas verbales se basaron en los sucesivos estudios de puntuación llevados a cabo a lo largo del desarrollo de la escala tanto para los ítems nuevos como para los procedentes de la versión anterior. En el capítulo 5 de este manual se ofrecen más detalles acerca de los estudios realizados sobre la puntuación de las pruebas de Comprensión verbal que requieren cierta corrección subjetiva de las respuestas verbales.

Baremos

Varios estudios indican que los baremos más antiguos producen puntuaciones infladas en las medidas de inteligencia (Flynn, 1999, 2007; Flynn y Weiss, 2007). Las puntuaciones de los tests deben basarse en información normativa actual y representativa de la población pertinente. Una completa descripción de la muestra española se presenta en el capítulo 4 de este manual.

Fiabilidad y validez

Se realizaron diversos estudios para demostrar la fiabilidad y validez de la escala. Los datos de retest se obtuvieron para todo el rango de edad, así como para las dos franjas de edad (2:6-3:11, 4:0-7:7). Mediante estudios de correlación se obtuvieron resultados de validez convergente y discriminante con los siguientes instrumentos: WPPSI-III (datos españoles); *Bayley Scales of Infant and Toddler Development, Third Edition* (Bayley-III; Bayley, 2005); WISC-IV; *Naglieri Nonverbal Ability Test-Second Edition* (NNAT2; Naglieri, 2007); NEPSY-II (Korkman, Kirk y Kemp, 2007); y *Behavior Assessment System for Children-Second Edition* (BASC-2 PRS; Reynolds y Kamphaus, 2004) en su forma para los padres. Estos estudios superan en número a los realizados para la WPPSI-III.

Durante la tipificación estadounidense de la WPPSI-IV también se recogieron datos de grupos especiales. La validez de constructo se obtuvo de diversos estudios de análisis factorial confirmatorio y de la comparación de las puntuaciones medias obtenidas por los grupos especiales y los grupos control. En los capítulos 5 y 6, respectivamente, se describen y se presentan los resultados de los estudios de fiabilidad y validez.

Suelo y techo de las pruebas

Se prestó una atención particular al efecto suelo y techo de cada prueba para garantizar que el rango de puntuaciones fuera representativo de los distintos niveles de aptitud cognitiva, desde el muy bajo ($40 \leq CIT \leq 69$) hasta el muy superior ($130 \leq CIT \leq 160$). A fin de medir adecuadamente el rango de aptitud, se añadieron ítems de poca y mucha dificultad a aquellas pruebas procedentes de versiones anteriores. Por ejemplo, se añadieron ítems gráficos para ampliar el suelo de las pruebas *Semejanzas* y *Comprensión*. Al incluir estos ítems gráficos, se amplió el rango de

puntuaciones escalares. Para el grupo de 4:0 a 4:2 años, el rango de puntuación escalar es de 1-19 para *Semejanzas* (en la WPPSI-III era de 6-19) y de 2-19 para *Comprensión* (en la WPPSI-III era de 5-19). Asimismo se mejoró el suelo de las pruebas nuevas respecto a las pruebas a las que sustituyeron. Por ejemplo, el rango de puntuación escalar de *Búsqueda de animales* y *Cancelación* es de 1-19 para el grupo de edad de 4:0 a 4:2 años (en la WPPSI-III era de 6-19 para *Búsqueda de símbolos* y de 5-19 para *Cancelación*). También se mejoró el suelo de *Clave de figuras* respecto a la prueba *Claves*, su equivalente en la WPPSI-III, y el rango de puntuación escalar actual para el grupo de edad de 4:0 a 4:2 años es de 3-19. Además, se añadieron ítems de mayor dificultad para ampliar el techo de *Vocabulario*, *Nombres*, *Cubos*, *Rompecabezas* y *Matrices*.

La utilidad clínica de los suelos y techos que se ampliaron se validó utilizando los datos obtenidos de niños con un diagnóstico de discapacidad intelectual leve o moderada, retraso cognitivo del desarrollo y alta capacidad intelectual. En el capítulo 6 se ofrece información detallada sobre estos y otros estudios con grupos especiales. Estas mejoras en el suelo y el techo de las pruebas permiten discriminar mejor el rendimiento de los niños situados en los puntos extremos del rango de aptitudes cognitivas; además, ofrecen un mayor rango de puntuación para el CI total correspondiente a los niños de 2:6 a 3:11 años: 40-160 (en comparación con el que ofrecía la WPPSI-III: 41-155).

Sesgo de los ítems

En la selección de los ítems de la WPPSI-IV se usaron las técnicas actuales destinadas a evitar los sesgos de los ítems. Por ello se han eliminado los ítems problemáticos según la opinión de los expertos y los datos empíricos de los análisis estadísticos sobre el sesgo. En todas las fases de desarrollo de la escala, expertos en evaluación de la inteligencia y en investigaciones transculturales revisaron los ítems basándose en el análisis de los posibles sesgos, la obsolescencia cultural, la relevancia del contenido y la utilidad clínica.

Además de todas estas revisiones, se usaron los datos empíricos de análisis de sesgo para servir de base en la toma de decisiones. Los resultados de los tradicionales análisis de sesgo de Mantel-Haenszel (Holland y Thayer, 1988) y de los análisis de sesgo de la teoría de respuesta al ítem (IRT; Hambleton, 1993) permitieron obtener información estadística sobre ítems potencialmente problemáticos.

Opciones del nivel de significación de los valores críticos

Las opciones del nivel de significación de los valores críticos, usadas para comparar puntuaciones, se ampliaron respecto a las de la WPPSI-III. El número de opciones pasó de dos a cuatro: la WPPSI-IV ofrece los niveles .01 y .10, además de los niveles .05 y .15 que presentaba la WPPSI-III. El nivel .01 es una opción aún más estricta que el nivel .05. Este cambio concede a los profesionales más flexibilidad en su intento de lograr un equilibrio entre la potencia estadística y el posible error de tipo I (falsos positivos) al comparar las diferencias. El nivel de significación puede ajustarse en función del número de comparaciones que hacer, y si tales comparaciones se han establecido a priori o se han realizado post hoc. En el capítulo 2 del *Manual de aplicación y corrección* y en el capítulo 5 de este manual se ofrecen más detalles sobre la adecuada selección de los niveles de significación de los valores críticos.

Incremento de la utilidad clínica

Todos los objetivos de la revisión previamente comentados guardan relación con la utilidad clínica de la escala. Por ejemplo, la mayor adecuación al desarrollo incrementa la validez de los resultados obtenidos. En los apartados siguientes se describen las características de la WPPSI-IV que mejoran la utilidad clínica.

Estructura de la escala

La estructura de la escala, expuesta antes en este capítulo, es novedosa y fue influenciada por los modelos estructurales de la inteligencia contemporáneos, la teoría del neurodesarrollo, la investigación sobre la neurocognición, y la utilidad clínica. Los estudios recientemente publicados de

análisis factorial de la WISC-IV y la WAIS-IV se basan en el modelo de cinco factores para ambas escalas (Benson et al., 2010; Keith et al., 2006; Weiss, Keith, Zhu y Chen, 2013; Weiss, Keith, Zhu y Chen, 2013) de forma similar a la estructura de la WPPSI-IV. En el capítulo 6 de este manual se presentan estos estudios, así como los resultados de los estudios de análisis factorial confirmatorio de la WPPSI-IV.

Las decisiones adoptadas sobre la estructura de la escala se fundamentan en los modelos neurocognitivos y del neurodesarrollo y en la utilidad clínica. El hecho de disponer de puntuaciones independientes para Visoespacial y Razonamiento fluido en el grupo de edad de 4:0 a 7:7 años incrementa la precisión de la interpretación. La incorporación de pruebas de memoria de trabajo mejora la utilidad clínica de la escala porque son sensibles a numerosos trastornos clínicos y medidas importantes de la evolución del niño (p. ej., rendimiento académico).

Los dominios cognitivos representados por la WPPSI-IV mediante las puntuaciones de los índices primarios han demostrado su utilidad en evaluaciones psicológicas de distinto tipo. El rendimiento en comprensión verbal es más bajo en los niños con trastorno de la lectura, trastorno del cálculo, trastornos del lenguaje y, en algunos estudios, trastorno autista de bajo funcionamiento (Chilosi et al., 2009; Compton, Fuchs, Fuchs, Lambert y Hamlett, 2012; Osmon, Smerz, Braun y Plambeck, 2006), pero generalmente no es un punto débil en los niños con trastornos del espectro autista de alto funcionamiento o trastorno de Asperger, a excepción de la prueba *Comprensión* (Mayes y Calhoun, 2008; Wechsler, 2003). Los déficits de la aptitud visoespacial se asocian con el trastorno del cálculo; en cambio, para los niños con trastornos del espectro autista, esta aptitud es un relativo punto fuerte, según indican diversos estudios (Mayes y Calhoun, 2008; Osmon et al., 2006; Wechsler, 2003, 2008). Los déficits del razonamiento fluido se relacionan con el trastorno de la comprensión lectora y el trastorno del cálculo (Fuchs et al., 2012; Osmon et al., 2006), pero no con el TDAH ni los trastornos del espectro autista de alto funcionamiento (Oliveras-Rentas, Kenworthy, Roberson, Martin y Wallace, 2012). La memoria de trabajo y la velocidad de procesamiento guardan más relación con diversos trastornos clínicos. Por ejemplo, la memoria de trabajo es sensible a los trastornos del aprendizaje, el TDAH, los trastornos de la lectura y los trastornos del espectro autista (Archibald y Gathercole, 2006a; Belleville, Ménard, Mottron y Ménard, 2006; Compton et al., 2012; Mayes y Calhoun, 2008; Osmon et al., 2006; Passolunghi, 2006; Pickering, 2006; Roodenrys, 2006; Swanson, 2006). La velocidad de procesamiento es sensible a los trastornos del aprendizaje, el TDAH y los trastornos del espectro autista, entre otros (Compton et al., 2012; Mayes y Calhoun, 2007, 2008). Cabe destacar que los niños con alta capacidad intelectual reflejan sus puntos fuertes en comprensión verbal, aptitud visoespacial y razonamiento fluido; sin embargo, en promedio, muestran un rendimiento más bajo en memoria de trabajo y velocidad de procesamiento (Raiford, Weiss, Rolfhus y Coalson, 2005; Rimm, Gilman y Silverman, 2008).

Aunque los tests de inteligencia no estén diseñados para su uso como medidas neuropsicológicas, las puntuaciones de los índices primarios de la WPPSI-IV representan procesos cognitivos importantes en la evaluación neuropsicológica. Los resultados de la WPPSI-IV pueden utilizarse para generar hipótesis sobre los déficits del procesamiento neuropsicológico, y pueden conceptualizarse e interpretarse desde el punto de vista neuropsicológico (Hale y Fiorello, 2004; Miller y Maricle, 2012). Por ejemplo, el ICV es, en parte, una medida de la expresión oral y del lenguaje receptivo. El IVE y el IVP se corresponden con los procesos cognitivos generalmente evaluados en los dominios visoespacial, y de velocidad y eficiencia. El IRF puede conceptualizarse como la medición de la resolución de problemas, la flexibilidad cognitiva y el razonamiento, en el dominio de la función ejecutiva. El IMT se diseñó para medir la memoria de trabajo visual, un proceso generalmente evaluado como parte del aprendizaje y la memoria (Miller y Maricle, 2012). Como ya se ha mencionado, estas aptitudes específicas parecen asociadas con la activación o el desarrollo de diferentes regiones cerebrales.

Metodología de comparación de las diferencias entre puntuaciones

El enfoque de la WPPSI-IV hacia la comparación de las diferencias entre puntuaciones es un cambio significativo respecto al de la WPPSI-III. Esta nueva metodología de la que seguidamente se ofrece un resumen, se expone con detalle en el capítulo 5 de este manual. Los puntos fuertes y débiles en los distintos índices primarios pueden evaluarse utilizando una estimación de la aptitud

general (la media de los índices primarios o el CI total) como puntuación de comparación. Los aspectos más globales de la evaluación del rendimiento (comparaciones con estimaciones generales del rendimiento) orientan los posteriores análisis del rendimiento en dominios más específicos (comparaciones entre dos índices primarios o dos pruebas). Este enfoque evaluativo, que va de las comparaciones generales a las específicas, puede reducir el número de comparaciones necesarias, de lo que resulta una menor probabilidad de errores de tipo I, es decir, falsos positivos (Grégoire, Coalson y Zhu, 2011; Longman, 2004; Naglieri y Paolitto, 2005). Las otras opciones del nivel de significación de los valores críticos también le ofrecen al profesional más control sobre los errores de tipo I, en función del motivo de consulta específico del niño.

Ampliación del rango de edad

El rango de edad de la WPPSI-IV se amplió hasta los 7 años 7 meses (en la WPPSI-III, 7 años 3 meses). Los profesionales consideraban conveniente esta ampliación del rango porque, de este modo, dispondrían de más flexibilidad al seleccionar un instrumento con el que evaluar a los niños de 7 años que iniciaran el segundo curso de primaria con retraso en el aprendizaje. Por consiguiente, se añadió un grupo de edad de 4 meses en el extremo superior del rango, de forma que este nuevo grupo se solapara con el mismo grupo de edad de la WISC-IV. Aunque el límite 7:7 puede parecer extraño, se seleccionó porque coincidía con el grupo correspondiente de la WISC-IV, que termina a los 7:7 años. De este modo, el rango de edad de la WPPSI-IV y el de la WISC-IV se solapan para los niños de 6:0 a 7:7 años. Por ello, los profesionales pueden elegir la medida que consideren adecuada para los niños de estas edades. En el capítulo 1 del *Manual de aplicación y corrección* se exponen los factores que deben tenerse en cuenta al elegir entre una medida u la otra.

Requisitos de lenguaje expresivo

En la WPPSI-IV se redujeron los requisitos de lenguaje expresivo respecto a los de la WPPSI-III. Se añadieron nuevos ítems a *Semejanzas* y *Comprensión* a fin de que los niños de 4:0 a 7:7 años con dificultades expresivas pudieran responder a algunos ítems aunque no pudieran hacerlo oralmente. Estos ítems, por tanto, permiten a estos niños demostrar su formación de conceptos verbales y su comprensión sin el requisito de dar una respuesta verbal. Solo *Vocabulario* y *Nombres* requieren respuestas expresivas en los ítems de baja dificultad, y ninguna de estas pruebas son principales para obtener los índices primarios ni el CI total en ninguna de las dos franjas de edad.

Dos índices secundarios ofrecen información sobre el rendimiento de los niños con problemas del lenguaje expresivo. El primero, el IAV, se obtiene a partir de las puntuaciones de *Dibujos* y *Nombres* y proporciona una estimación del nivel de adquisición del vocabulario. La comparación de las puntuaciones de estas dos pruebas puede ofrecer información sobre las destrezas del lenguaje receptivo y expresivo. El segundo índice secundario, el INV, se obtiene a partir de pruebas que no requieren ninguna respuesta expresiva y que pertenecen a todos los dominios cognitivos de la WPPSI-IV, a excepción de comprensión verbal. El INV proporciona una estimación más adecuada de la aptitud general de los niños con retrasos del lenguaje expresivo significativos o con otros trastornos clínicos asociados a dificultades verbales expresivas (p. ej., trastornos del espectro autista). Además, ofrece una estimación útil de la aptitud general de los niños sordos o con dificultades de audición, así como de los niños que están aprendiendo español. Este índice abarca más de un dominio cognitivo y, por lo tanto, proporciona una estimación más sólida de la aptitud general que los índices primarios, como el IVE o el IRF. En el anexo B de este manual se ofrece más información sobre los usos posibles del IAV y el INV, así como las tablas necesarias para obtener estas puntuaciones secundarias.

Puntuaciones de los índices secundarios

Además del IAV y el INV, la WPPSI-IV ofrece el ICG y el ICC. El ICG es una puntuación compuesta que se basa en las pruebas que contribuyen al CI total de Comprensión verbal y Visoespacial para los niños más pequeños (de 2:6 a 3:11 años) y de Comprensión verbal, Visoespacial y Razonamiento fluido para los niños más mayores (de 4:0 a 7:7 años). Conceptualmente,

el ICG ofrece una estimación de la aptitud intelectual general menos dependiente de la memoria de trabajo y de la velocidad de procesamiento que el CI total.

El ICC es una puntuación compuesta que se basa en las pruebas que contribuyen al IMT y al IVP. Conceptualmente, el ICC ofrece una estimación de la eficiencia con la que se procesa la información cognitiva durante el aprendizaje, la resolución de problemas y el razonamiento de nivel superior. El ICC solo puede obtenerse para los niños en la franja de edad superior (de 4:0 a 7:7 años). En el anexo C de este manual se describen estas puntuaciones y sus aplicaciones.

Grupos especiales

Para aumentar la utilidad clínica de la escala, durante el proceso de tipificación estadounidense se realizaron diversos estudios con 12 grupos especiales, entre los que pueden citarse: alta capacidad intelectual, discapacidad intelectual leve, discapacidad intelectual moderada, retraso del desarrollo cognitivo, factores de riesgo para el desarrollo, problemas previos a la lectoescritura, TDAH, conducta disruptiva, trastorno del lenguaje expresivo, trastorno mixto del lenguaje receptivo-expresivo, trastorno autista y trastorno de Asperger. Los estudios con los grupos clínicos se describen en el capítulo 6 de este manual.

Desarrollo de la versión estadounidense de la WPPSI-IV

El proceso de investigación previo a la publicación de la WPPSI-IV tuvo una duración de cuatro años, y cada una de sus fases conllevó mejoras en la escala. En este capítulo se describe todo el proceso, desde los procedimientos y fases de investigación hasta los criterios de calidad adoptados. Para finalizar, se expone detalladamente la fase de tipificación y la composición final de la escala.

Directrices de investigación

Los *Standards for Educational and Psychological Testing* (*Standards*; American Educational Research Association, American Psychological Association y National Council on Measurement in Education [AERA], 1999) han sido recursos básicos durante todo el proyecto de investigación de la WPPSI-IV y han aportado criterios para «el valor de los tests, las buenas prácticas en la evaluación y las consecuencias del uso de los tests». En *Standards* figura abundante información sobre estos temas y otros aspectos relacionados con el desarrollo y la evaluación de los tests.

Principales fases de la investigación

Antes de establecer la estructura final de la escala, se siguieron varias etapas experimentales, basadas en un minucioso proyecto de trabajo. La WPPSI-IV se desarrolló en cinco fases: desarrollo conceptual, versión piloto, experimentación, tipificación, y composición y evaluación final. En los siguientes apartados se presentan brevemente estas fases, prestando especial atención a los estudios previos y a la metodología seguida.

Desarrollo conceptual

Esta primera fase se centró en definir los objetivos de la revisión de la escala, así como la metodología que se seguiría para lograr dichos objetivos. Se consultaron numerosas fuentes bibliográficas y los estudios de las encuestas realizadas a los profesionales y usuarios, así como estudios de expertos y examinadores de Pearson a fin de establecer una primera serie de objetivos. Posteriormente se formó el equipo del proyecto, compuesto por el equipo investigador (equipo de desarrollo), el director del producto, el director del proyecto, los consultores de Pearson y el equipo asesor para la WPPSI-IV. Asimismo participaron en el proyecto diversos miembros de Pearson especialistas en edición, producción, investigación de campo (recolección de datos) y psicometría. Y, finalmente, se contó con la inestimable colaboración de expertos, internos y externos, en psicometría, en las escalas de inteligencia de Wechsler, así como en otras áreas de investigación relacionadas y en el desarrollo de tests.

Equipo asesor

El equipo investigador contó, a lo largo del proyecto, con la colaboración de reconocidos expertos en neuropsicología clínica y desarrollo cognitivo de los niños. Tras cada una de las principales fases (p. ej., piloto, experimentación o tipificación) se revisó el plan de trabajo de la investigación teniendo en cuenta los resultados psicométricos obtenidos y las opiniones del equipo del proyecto (incluidos los asesores), expertos, revisores de sesgos y examinadores.

Investigación profesional

En los inicios del proceso de desarrollo de la WPPSI-IV se llevaron a cabo dos encuestas en línea (N = 300 y N = 77) con profesionales que utilizaban la WPPSI-III y profesionales con experiencia en la evaluación de la primera infancia. El equipo del proyecto revisó las aportaciones de dichos profesionales. Posteriormente se realizó una encuesta telefónica de seguimiento (N = 15) y se analizaron las respuestas que se centraban en aspectos más específicos de los patrones de uso, las prioridades y los requisitos de la amplitud del contenido de la escala. A partir de todos los datos recopilados, se elaboró el plan de trabajo y se desarrolló la versión de la escala para la fase piloto.

Estudios semiestructurados de expertos y examinadores

En todas las etapas de recogida de datos se realizaron estudios semiestructurados que permitieron a los expertos y examinadores valorar y comentar las versiones de la escala en términos de adecuación al desarrollo, utilidad clínica y facilidad de uso. Los resultados de estos estudios se resumieron y analizaron con el equipo del proyecto. Asimismo, a lo largo de las fases de desarrollo de la escala, se realizaron diversas modificaciones del plan de trabajo y de las distintas versiones de la escala a partir de las opiniones de los expertos y los examinadores y los datos del análisis cualitativo.

Estudios piloto

El principal objetivo de esta fase fue crear la versión de la escala que se usaría en la fase de experimentación. Se realizó una serie de tres estudios minipiloto (N = 10, 11 y 23) y dos estudios piloto (N = 364 y 337). En cada uno de ellos se utilizó una versión de la escala formada por grupos de pruebas ya existentes en la WPPSI-III y pruebas nuevas; todas ellas eran experimentales y se estaba valorando su inclusión en la versión de experimentación. Esta fase se centró en el análisis del contenido y la relevancia de los ítems, la adecuación de los techos y suelos de las pruebas, la claridad de las consignas tanto para el examinador como para el niño, la identificación de los procesos de respuesta, los procedimientos de aplicación, los criterios de puntuación, el sesgo de los ítems y otros aspectos psicométricos relevantes.

Experimentación

En la fase de experimentación se utilizó una versión de la escala formada por 18 pruebas. Los datos se obtuvieron de una muestra estratificada de 437 niños representativa de la población de Estados Unidos. El análisis de los datos de la Oficina del Censo de Estados Unidos (2008) permitió obtener la base para la estratificación de los distintos grupos de edad según las variables siguientes: edad, sexo, nivel educativo de los padres y zona geográfica. Dentro de cada grupo de edad, la muestra de experimentación se aproximó mucho a los datos del censo para las cuatro variables demográficas.

Esta muestra, más amplia y representativa, permitió volver a examinar algunas de las cuestiones planteadas en la fase piloto y resolver cuestiones nuevas. El orden de los ítems se refinó teniendo en cuenta estimaciones más precisas de la dificultad relativa de los ítems, y se llevaron a acabo análisis factoriales exploratorios y confirmatorios para determinar la estructura factorial de la escala. Además, se recopilaron datos de grupos especiales (niños con un diagnóstico de alta capacidad intelectual [N = 19], discapacidad intelectual [N = 27] y TDAH [N = 31]) para analizar la adecuación de los suelos y techos de las pruebas y la utilidad clínica de la escala.

Se revisaron los datos de experimentación, así como los resultados sobre la adecuación al desarrollo, la facilidad de uso, la utilidad clínica y las propiedades psicométricas de esta versión de la escala. De este análisis se concluyó que las instrucciones y las consignas, los requisitos de las tareas, el tiempo de aplicación, la fiabilidad, el contenido y orden de los ítems, los techos y suelos de las pruebas, la estructura factorial y la utilidad clínica de la versión de experimentación tenían la calidad suficiente para conservarlos en la fase de tipificación.

Tipificación

Tomando en cuenta los resultados de los estudios piloto y de experimentación, se preparó la versión de tipificación de la WPPSI-IV. Dada la calidad de los resultados obtenidos con la versión de experimentación, para la versión de tipificación se mantuvieron sin cambios los ítems y su orden, los procedimientos de aplicación, las instrucciones para los examinadores, las consignas verbales para los niños, los materiales manipulativos y los materiales que conforman los estímulos.

La fase de tipificación se centró en la obtención de normas y la comprobación de la fiabilidad, validez y utilidad clínica de la escala. Los datos se obtuvieron de una muestra estratificada de 1700 niños de edades comprendidas entre los 2:6 y 7:7 años, así como de muestras de niños pertenecientes a 13 grupos especiales. Para demostrar la validez de la escala, se aplicó la WPPSI-IV y otras medidas (p. ej., WPPSI-III, Bayley-III, y WISC-IV) a otras muestras de niños. Las características de las muestras que participaron en los estudios de validez concurrente y de grupos especiales figuran en el capítulo 6 de este manual.

El equipo investigador analizó los resultados psicométricos de los estudios de tipificación junto con los informes realizados por expertos y examinadores. A partir de todos los datos recopilados a lo largo del proceso de investigación, se estableció la estructura final de la escala para componer y evaluar la versión definitiva.

Composición final y evaluación de la WPPSI-IV

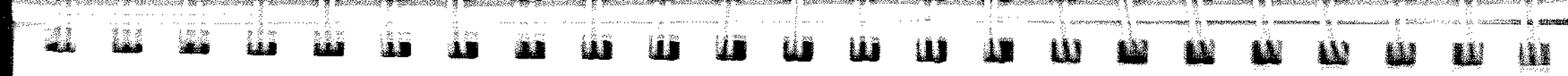
Consistencia de los ítems, las consignas y los estímulos

A fin de garantizar que se conservaban los mejores ítems tras la selección efectuada, las versiones experimentales de la WPPSI-IV incluían más ítems de los que serían necesarios para las pruebas de la versión definitiva. A lo largo de las diversas fases de desarrollo de la escala, se evaluaron los ítems con el objeto de conservarlos, modificarlos o eliminarlos. Por lo tanto, todos los ítems habían sido evaluados en diversos momentos antes de la tipificación de la escala. La selección final de los ítems se realizó tras analizar los datos obtenidos de las muestras de tipificación y para eliminar los ítems redundantes y los ítems que eran demasiado fáciles o demasiado difíciles. Siguiendo este criterio, se eliminaron como promedio cinco ítems de cada prueba, y, en la medida de lo posible, se mantuvo el orden de los ítems dentro de las pruebas. Los ítems establecidos como puntos de comienzo en la versión de tipificación se mantuvieron como tales en la versión final.

Asimismo, en la versión final de la escala, no se modificaron las consignas para el niño, ni los ítems del cuaderno de estímulos. Aunque se modificó ligeramente la formulación de las instrucciones para el examinador a fin de que fueran más claras, los cambios realizados no suponen ninguna alteración de los procedimientos de aplicación estandarizados de la escala.

Consistencia del orden de aplicación

El desarrollo de la estructura final de la escala incluyó el establecimiento del tipo de prueba (principal, complementaria u opcional), lo que dio lugar a cambios menores en el orden de aplicación de las pruebas para los niños de 4:0 a 7:7 años. Estos cambios de orden se realizaron para que primero se aplicaran las pruebas principales que permiten obtener el CI total, y después las pruebas restantes de los índices primarios y secundarios, respectivamente. Y para evaluar los posibles efectos de este nuevo orden de aplicación de las pruebas se llevó a cabo un estudio especial utilizando un diseño intersujeto. Se aplicaron las 15 pruebas en el orden final propuesto a 44 niños de 4:0 a 7:7 años. Las puntuaciones de las pruebas y las puntuaciones compuestas de este grupo se compararon con las del grupo control, formado por 44 niños de la muestra de tipificación. Las comparaciones no indicaron diferencias significativas en las puntuaciones medias de estos grupos.



Adaptación española de la WPPSI-IV

El objetivo de la adaptación española de la WPPSI-IV ha sido proporcionar a los profesionales españoles una herramienta adaptada a sus necesidades prácticas y clínicas, y unos baremos actualizados, acordes con los cambios poblacionales españoles.

La adaptación española de la WPPSI-IV se inició en septiembre del 2012. A lo largo de estos meses no se ha realizado una simple traducción, sino una adaptación completa de la escala. Se ha llevado a cabo un desarrollo técnico y psicométrico completo que permite asegurar la validez y fiabilidad de la escala en la cultura española. Como es bien sabido, las diferencias culturales, lingüísticas y educativas entre países generan a menudo representaciones conceptuales diferentes de un mismo término. Por ello, la adaptación de la WPPSI-IV, especialmente en el caso de las pruebas verbales, ha comportado una transposición cultural. El objetivo de una transposición cultural es encontrar las equivalencias entre las culturas que permitan adaptar la prueba conservando su esencia. Con este objetivo, en el proceso de adaptación española, y a partir del análisis del contenido original, se tuvieron en cuenta diversos criterios para seleccionar los ítems, redactar las consignas de aplicación y establecer los criterios de puntuación.

En este capítulo se presentan, en primer lugar, los procedimientos empleados para seleccionar los ítems de la versión experimental, las características de la muestra experimental, y los análisis que permitieron determinar el orden que debía ocupar cada ítem dentro de cada prueba. En segundo lugar, se presenta la información referente al proceso de tipificación, incluidas las modificaciones de los ítems respecto a la WPPSI-III, las características de la muestra de tipificación y su representatividad. Por último, se exponen los métodos adoptados para la creación de los baremos y para la obtención de las puntuaciones compuestas, entre otros.

Fase experimental

Las pruebas que componen la WPPSI-IV pueden clasificarse en función de múltiples criterios. Teniendo en cuenta su contenido, podemos diferenciar entre pruebas mayoritariamente verbales o pruebas con escaso contenido verbal. En este sentido, los ítems que tradicionalmente requieren ser adaptados son los de las pruebas verbales, puesto que con frecuencia, al traducir un ítem de un idioma a otro, se pierde la esencia o el significado original que pretendía.

La adaptación de cualquier instrumento de evaluación psicológica se inicia con el análisis minucioso del contenido de la versión original en la que se basa. La versión experimental española de la WPPSI-IV se apoyó en la versión original estadounidense, pero teniendo en cuenta la versión anterior de la escala para garantizar cierta continuidad (WPPSI-III; 2009). Los objetivos de la fase experimental fueron traducir, revisar y adaptar los ítems de las pruebas verbales, crear nuevos ítems equivalentes a los originales, probar psicométricamente el funcionamiento de los ítems y establecer el orden que debían ocupar en la versión de tipificación.

Selección de los ítems experimentales

En la versión experimental de la WPPSI-IV se incluyeron tanto los ítems verbales como los ítems gráficos de las pruebas con contenido verbal: *Información, Semejanzas, Vocabulario, Comprensión, Dibujos y Nombres.*

El proceso de elaboración de la versión experimental se llevó a cabo en varios pasos. En primer lugar, el equipo de I+D junto con una experta traductora y filóloga, tradujeron al castellano todos los ítems de las pruebas verbales de la versión original estadounidense. En segundo lugar, se examinaron en detalle tanto la traducción de los ítems verbales como la imagen que ilustraban los ítems gráficos. Se decidió conservar, modificar o eliminar cada ítem en función de la idoneidad de su contenido y la adecuación a la cultura española. En el caso de los ítems conservados o modificados, se tuvo en cuenta que la traducción y/o la imagen ilustrada, recogieran el significado y el objetivo primario de la versión original, así como que mantuvieran su equivalencia en cuanto a la dificultad. Finalmente, para reemplazar los ítems eliminados, se crearon ítems nuevos para cada prueba verbal.

La tabla 4.1 presenta un resumen de los ítems conservados, eliminados y nuevos en la versión experimental española, respecto a la versión original estadounidense. En la tabla puede observarse que la versión experimental contó con más ítems que los incluidos en la versión definitiva. Este procedimiento permitió, tras analizar los resultados de la muestra experimental, seleccionar los ítems con mejores propiedades psicométricas.

Tabla 4.1. Ítems conservados, eliminados y nuevos respecto a la versión estadounidense para la versión experimental española

	Versión estadounidense WPPSI-IV	Versión experimental española WPPSI-IV				
		Conservados				
Prueba	*Total ítems*	*Sin modificaciones*	*Con ligeras modificaciones*	*Eliminados*	*Nuevos*	*Total ítems experimentados*
Información	29	26	1	2	8	35
Semejanzas	23	19	2	2	9	30
Vocabulario	23	18	0	5	11	29
Comprensión	22	17	0	5	10	27
Dibujos	31	26	1	4	9	36
Nombres	24	21	1	2	9	31

Nota: Las cifras de la tabla representan la cantidad de ítems.

Al crear nuevos ítems para la versión experimental se tuvieron en cuenta los siguientes criterios:

- Ítems con nivel de dificultad equivalente a los ítems de la versión original estadounidense.
- Ítems con respuestas fácilmente clasificables en 0, 1 o 2 puntos, según la prueba.
- Ítems sin significado ambiguo o diferente en otras zonas hispanohablantes.
- Ítems con una formulación clara y precisa.
- Ítems sin connotaciones religiosas o peyorativas.
- Ítems sin sesgo de género.
- Ítems apropiados para todas las edades.

Además, se tuvo en cuenta que los ítems cubrieran una amplia variedad de conocimientos y temáticas. Así, en *Información* se procuró que la proporción de ítems de las áreas de ciencias sociales, ciencias naturales y cuerpo humano, entre otras, fuera equivalente. Con el objeto de evitar cualquier sobrerrepresentación en el formato de los ítems, en *Vocabulario* también se controló la frecuencia de aparición de verbos y sustantivos; y en Dibujos, la aparición de verbos, sustantivos y adjetivos. En *Comprensión* y *Semejanzas*, todos los ítems se formularon o reformularon de manera precisa para evitar respuestas muy concretas o personalizadas que no responden al ítem. Y en *Nombres* se tuvo en cuenta la proporción de animales y objetos ilustrados.

Finalmente, la versión experimental de la WPPSI-IV estuvo compuesta por 35 ítems para *Información*, 30 para *Semejanzas*, 29 para *Vocabulario*, 27 para *Comprensión*, 36 para *Dibujos* y 31 para *Nombres*. Para evaluar estos ítems, se diseñó una muestra experimental estratificada y representativa de los niños de la población española y se seleccionaron los examinadores que se encargarían de la recogida de datos.

Selección y formación de los examinadores

Con la colaboración de los colegios oficiales de psicólogos de toda España, se contactó con psicólogos infantiles que tenían amplia experiencia en evaluación de la inteligencia y de las aptitudes cognitivas y que procedían de diversos campos de la psicología (clínico, educativo, forense, etc.).

Los 20 psicólogos seleccionados, procedentes de distintas zonas geográficas de España, participaron en un curso de formación presencial donde se expusieron los objetivos y requerimientos de la fase experimental, los procedimientos específicos de aplicación de la WPPSI-IV y los cambios respecto a la versión anterior de la escala.

La formación permitió homogeneizar los procedimientos de aplicación entre examinadores y aclarar las dudas referentes a las instrucciones de aplicación. Se pidió a todos ellos que, durante cada aplicación, transcribiesen palabra por palabra las respuestas de los niños y que indicaran sus reacciones frente a los ítems y al material. La corrección y la puntuación de las respuestas fueron realizadas por el departamento de I+D de Pearson Clinical & Talent Assessment.

Aplicación de la versión experimental

Para la aplicación de la versión experimental, se proporcionó a los examinadores todo el material necesario y una plantilla personalizada de los niños que cada uno debía reclutar, atendiendo a los criterios de estratificación. El material experimental incluía un manual donde se especificaban las instrucciones y consignas de aplicación y los criterios de inclusión de la muestra; el cuadernillo de anotación para el registro de todas las respuestas del niño y sus características sociodemográficas; y el cuaderno de estímulos que contenía los ítems gráficos de *Información, Semejanzas, Vocabulario, Comprensión, Dibujos* y *Nombres*.

En la fase experimental de la WPPSI-IV no se aplicaron los límites de tiempo y reglas de comienzo, retorno y terminación de las pruebas. Se pidió a los examinadores que aplicaran todos los ítems con la intención de valorar, posteriormente, la dificultad y posición que debía ocupar cada ítem en cada prueba.

La aplicación de la forma experimental también permitió conocer las respuestas que con más frecuencia daban los niños de la muestra, y a partir de ellas se elaboraron los criterios de puntuación para la versión de tipificación.

Control y seguimiento

Para controlar y garantizar la calidad de los datos recogidos, se puso en marcha un sistema de evaluación «luz verde». Este proceso consistió en corregir y analizar minuciosamente el primer protocolo enviado por cada uno de los examinadores, para detectar cualquier tipo de error en la aplicación de las pruebas o en el registro de las respuestas. Si el protocolo cumplía los requisitos de calidad propuestos inicialmente por el departamento de I+D de Pearson, se daba al examinador «luz verde» para que prosiguiera sus aplicaciones. En caso contrario, se contactaba con el examinador para explicarle en detalle los errores u omisiones observados en su primer protocolo, y se le proporcionaban las indicaciones necesarias que debía tener en cuenta para sus próximas aplicaciones. El segundo protocolo enviado por el examinador era de nuevo evaluado cuidadosamente para comprobar la incorporación de las recomendaciones.

Para garantizar el cumplimiento de los plazos y la buena calidad de los datos recogidos, se pidió a los examinadores que enviaran al menos un protocolo semanalmente.

Muestra experimental

La recogida de datos de la fase experimental de la WPPSI-IV se realizó entre octubre del 2012 y enero del 2013 sobre una muestra de 189 niños, estratificada por edad, sexo, nivel educativo de los padres, tipo de población y zona geográfica, siguiendo las proporciones del último Censo de Población y Viviendas español (Instituto Nacional de Estadística; INE, 2001 [edición disponible en el momento de realizar este estudio]). Todos los padres de los niños que participaron en la fase experimental firmaron un documento de consentimiento informado.

Edad. La forma experimental se aplicó a cinco grupos de edad: 2:6-2:11 (n = 38), 3:6-3:11 (n = 37), 4:0-4:11 (n = 39), 5:6-6:6 (n = 36) y 7:0-7:6 (n = 39).

Sexo. La distribución por sexo fue equivalente en todos los grupos de edad. La muestra total estuvo compuesta por 95 niños (50.3 %) y 94 niñas (49.7 %).

Nivel educativo de los padres. Se clasificó el nivel educativo de los padres o tutores legales de los niños en cuatro categorías. En el caso de que ambos padres residieran con el niño y tuviesen distinto nivel educativo, se tuvo en cuenta el nivel educativo superior. La descripción de las categorías y el porcentaje de niños de la muestra experimental incluidos en cada categoría fueron los siguientes:

- Sin estudios: estudios primarios sin finalizar o inferior (3.7 %).
- Primer grado: graduado escolar (8.° de EGB) o 2.° de ESO o bachillerato elemental completados (11.6 %).
- Segundo grado: ESO o bachilllerato superior o formación profesional o COU completados (63.5 %).
- Tercer grado: diplomatura o licenciatura o titulaciones superiores completadas (21.2 %).

Tipo de población. También se consideró como una variable importante de estratificación el tipo de población donde residían los niños. Siguiendo la clasificación utilizada por el censo, se definieron tres categorías. La descripción de cada categoría y el porcentaje de niños de la muestra experimental evaluados en cada tipo de población fueron los siguientes:

- Urbana: población igual o superior a 50 000 habitantes (69.3 %).
- Intermedia: población entre 10 000 y 49 999 habitantes (16.9 %).
- Rural: población inferior a 10 000 habitantes (13.8 %).

Zona geográfica. Se determinaron cuatro zonas geográficas para que la muestra experimental fuese representativa de todo el territorio peninsular e insular. La descripción de las cuatro zonas y el porcentaje de niños de la muestra experimental evaluados en cada zona fueron los siguientes:

- Norte: Asturias, Cantabria, Castilla y León, Galicia, La Rioja, Navarra y País Vasco (24.3 %).
- Sur: Andalucía, Islas Canarias y Murcia (26.5 %).
- Este: Aragón, Cataluña, Comunidad Valenciana e Islas Baleares (28.0 %).
- Centro: Castilla-La Mancha, Extremadura y Madrid (21.2 %).

Análisis de los datos de la fase experimental

Los datos recogidos a partir de la muestra de la fase experimental fueron sometidos a análisis estadísticos. Solo los ítems que cumplieron con las cualidades psicométricas requeridas fueron seleccionados para la versión definitiva de la escala. En cada ítem se analizaron los siguientes aspectos:

- **Índice de dificultad.** El índice de dificultad de un ítem se define como el cociente entre el número de niños que han acertado el ítem y el número total de niños que han intentado resolver el ítem. El valor mínimo del índice de dificultad es 0

(ningún niño acierta el ítem) y el valor máximo es 1 (todos los niños que lo intentan lo aciertan). Si el índice de dificultad se acerca a 0, indica que el ítem ha resultado muy difícil; si se acerca a 1, indica que ha resultado muy fácil; y si se acerca a 0.5, indica una dificultad media. El índice de dificultad permite fijar el orden de presentación de los ítems para cada prueba.

- **El parámetro de dificultad según el modelo de Rasch.** El índice de dificultad tiene el inconveniente de ser dependiente de la muestra utilizada para calcularlo. Si los niños tienen un nivel de aptitud elevado, el ítem puede parecer fácil; si los niños tienen un nivel de aptitud bajo, el mismo ítem puede parecer difícil. La *teoría de la respuesta al ítem* (TRI) *constituye un nuevo enfoque en psicometría* que permite superar esta limitación calculando índices de dificultad invariantes respecto a la muestra. El modelo de Rasch es el índice más conocido de la TRI y permitió estimar el parámetro de dificultad de los ítems de la fase experimental.
- **Índice de discriminación.** Un buen ítem tiene que diferenciar, de la forma más clara posible, a los niños en función de su nivel de aptitud. Por ejemplo, los niños con un nivel de aptitud inferior a la media responderían incorrectamente a un ítem de dificultad mediana (0.5), mientras que los niños con un nivel de aptitud superior a la media responderían correctamente a ese mismo ítem. Si no es el caso, entonces ese ítem demuestra ser poco discriminativo. Para cada prueba, las correlaciones entre la puntuación en el ítem y la puntuación directa total se utilizaron como indicadores de la cualidad discriminante del ítem.

Los análisis estadísticos realizados sobre la muestra de la fase experimental permitieron seleccionar los ítems que formarían parte de la versión de tipificación española de la WPPSI-IV y asegurar sus cualidades métricas. A partir de esta versión se llevó a cabo el estudio de tipificación.

Fase de tipificación

Finalizada la fase experimental, un panel de expertos en evaluación y adaptación de instrumentos examinó las pruebas que por su escaso contenido verbal no habían sido probadas en la fase experimental, y dictaminó que solo eran necesarias pequeñas modificaciones en dos ítems de la prueba *Conceptos*, y que el resto de los ítems no mostraban sesgos aparentes. Esto es así porque, durante la fase de desarrollo original de la WPPSI-IV, el equipo de I+D estadounidense consultó con equipos de investigación del resto de países la adecuación del contenido de estos ítems en cada uno de sus países. La intención no era otra que la de elaborar una herramienta que fuera, en la medida de lo posible, culturalmente adecuada a la mayoría de los países, o que al menos requiriera mínimos cambios.

Modificaciones respecto a la WPPSI-III

Todas las mejoras efectuadas en la versión estadounidense de la WPPSI-IV en relación con su predecesora, la WPPSI-III, fueron aplicadas a la versión española de la WPPSI-IV: se eliminaron las pruebas *Adivinanzas, Búsqueda de símbolos, Claves* y *Figuras incompletas*; se añadieron cinco pruebas nuevas *Reconocimiento, Localización, Búsqueda de animales, Cancelación* y *Clave de figuras*; se actualizaron los ítems; y se simplificaron las consignas y los procedimientos de aplicación. La tabla 4.2 resume los principales cambios realizados en la adaptación española de la WPPSI-IV, respecto a la adaptación española de la WPPSI-III.

Tabla 4.2. Cambios en la adaptación española de la WPPSI-IV respecto a la WPPSI-III

	Adaptación española WPPSI-III	Adaptación española WPPSI-IV			
Prueba	**Total ítems**	**Conservados***	**Eliminados**	**Nuevos**	**Total ítems**
Cubos	20	13	7	4	17
Información	34	14	20	15	29
Matrices	29	21	8	5	26
Búsqueda de animales	–	–	–	66	66
Reconocimiento	–	–	–	35	35
Semejanzas	24	17	7	6	23
Conceptos	28	22	6	5	27
Cancelación	–	–	–	2	2
Localización	–	–	–	20	20
Rompecabezas	14	11	3	2	13
Vocabulario	25	5	20	18	23
Clave de figuras	–	–	–	72	72
Comprensión	20	9	11	13	22
Dibujos	38	21	17	10	31
Nombres	30	17	13	7	24

Nota: Las cifras de la tabla representan la cantidad de ítems.

* Algunos de los ítems conservados fueron ligeramente modificados.

Diseño de la muestra de tipificación

Para diseñar una muestra lo más representativa posible de la población española, se siguieron las proporciones y los datos del Censo de Población y Viviendas español (INE, 2001), que era el que aportaba todos los datos referentes al cruce de las variables de estratificación. La muestra de tipificación se planeó sobre una muestra de 900 niños, estratificada según las variables demográficas siguientes: edad, sexo, nivel educativo, tipo de población y zona geográfica.

Edad. La muestra se diseñó teniendo en cuenta 9 grupos de edad, de 100 niños cada uno: 2:6-2:11; 3:0-3:5, 3:6-3:11; 4:0-4:5, 4:6-4:11; 5:0-5:5, 5:6-5:11, 6:0-6:11 y 7:0-7:7.

Sexo. Porcentaje equivalente de niños y niñas.

Nivel educativo de los padres. Se determinó el nivel educativo de los padres o tutores legales de los niños en función de cuatro categorías. En el caso de que ambos padres residieran con el niño y tuviesen distinto nivel educativo, se tuvo en cuenta el nivel educativo superior. Las cuatro categorías utilizadas para estratificar la muestra fueron:

- Sin estudios: estudios primarios sin finalizar o inferior.
- Primer grado: graduado escolar (8.º de EGB) o 2.º de ESO o bachillerato elemental completados.
- Segundo grado: ESO o bachilllerato superior o formación profesional o COU completados.
- Tercer grado: diplomatura o licenciatura o titulaciones superiores completadas.

Tipo de población. También se consideró como una variable importante de estratificación el tipo de población donde residían los niños. Siguiendo la clasificación utilizada por el censo, se definieron tres categorías:

- Urbana: población igual o superior a 50 000 habitantes.
- Intermedia: población entre 10 000 y 49 999 habitantes.
- Rural: población inferior a 10 000 habitantes.

Zona geográfica. A fin de que los baremos estuvieran construidos a partir de una muestra muy diversa, procedente de distintas zonas geográficas de nuestro país, se planificó la recogida de datos en todo el territorio peninsular e insular. Se determinaron cuatro zonas geográficas:

- Norte: Asturias, Cantabria, Castilla y León, Galicia, La Rioja, Navarra y País Vasco.
- Sur: Andalucía, Ceuta, Islas Canarias, Melilla y Murcia.
- Este: Aragón, Cataluña, Comunidad Valenciana e Islas Baleares.
- Centro: Castilla-La Mancha, Extremadura y Madrid.

Paralelamente al proyecto de tipificación, se diseñó la muestra para los estudios de fiabilidad test-retest y de validez de la WPPSI-IV con otras medidas (WPPSI-III). Se recogieron varias muestras independientes estratificadas por edad, sexo y nivel educativo de los padres. Las características de estas muestras y los resultados de estos estudios pueden consultarse en el capítulo 6 de este manual.

El plan de muestreo presentado en este apartado permitió elaborar las tablas de estratificación personalizadas para cada examinador.

Selección y formación de los examinadores

Siguiendo los mismos criterios que en la fase experimental, se seleccionó a 52 examinadores para la fase de tipificación. Ocho examinadores que habían participado en la fase experimental de la WPPSI-IV participaron también en la fase de tipificación.

Se realizó un curso de formación para todos los examinadores. Durante la formación se expusieron los objetivos y los requerimientos de la fase de tipificación, así como una explicación detallada acerca del procedimiento específico de aplicación de la WPPSI-IV en esta fase y los cambios respecto a la versión anterior.

La formación permitió homogeneizar los procedimientos de aplicación entre examinadores y aclarar las dudas referentes a las instrucciones de aplicación. Se pidió a todos los examinadores que durante las aplicaciones registraran todas las respuestas del niño, y observaciones relevantes, a fin de facilitar la labor de puntuación realizada por el equipo de corrección de Pearson Clinical & Talent Assessment.

Aplicación de la versión de tipificación

La fase de tipificación se inició en marzo del 2013 y concluyó en noviembre del 2013.

Para la aplicación de la forma de tipificación se proporcionó a todos los examinadores todo el material necesario y una plantilla personalizada de los niños que cada uno debía reclutar, atendiendo a los criterios de estratificación. El material de tipificación incluía un manual donde se especificaban las instrucciones, las consignas de aplicación y los criterios de inclusión de la muestra; dos cuadernillos de anotación, en función de la edad, para el registro de las respuestas del niño y sus características sociodemográficas; tres cuadernos de estímulos con los ítems gráficos de las pruebas; tres cuadernillos de respuesta y un tampón para las pruebas *Búsqueda de animales, Cancelación* y *Clave de figuras*; y todo el material manipulativo necesario para aplicar *Localización, Rompecabezas* y *Cubos*. Estos materiales fueron iguales a los definitivos.

En esta fase, y a diferencia de la fase experimental, se incluyeron las limitaciones de tiempo, las reglas de comienzo, retorno y terminación. Se pidió a los examinadores que aplicaran la escala siguiendo de manera estricta las reglas y los procedimientos definitivos, a fin de obtener resultados válidos para la creación de los baremos.

Puntuación de los protocolos

Para controlar y garantizar la calidad de los datos recogidos, se puso en marcha el mismo sistema de evaluación «luz verde» adoptado en la fase experimental.

Para la corrección de las pruebas verbales de la fase de tipificación se partió de las respuestas más frecuentes que los niños dieron a los ítems en la fase experimental. A fin de crear los criterios de puntuación para las pruebas verbales, y diferenciar bien las respuestas en 0, 1 y 2 puntos, se analizaron las respuestas más frecuentes que los niños dieron a los ítems en la fase experimental. Esta labor se realizó con la ayuda de una filóloga experta y en función de las definiciones recogidas en distintos diccionarios de lengua castellana. Los criterios de puntuación para el resto de las pruebas no requirieron ajustes importantes puesto que son en su mayoría muy objetivas. La puntuación de los protocolos fue realizada por el equipo de correctores del departamento de I+D de Pearson Clinical & Talent Assessment, previamente formado y entrenado para esta tarea. Aunque el sistema de corrección de las pruebas de la WPPSI-IV no implica una dificultad elevada, el control de la fiabilidad de la puntuación es fundamental. Por ello, se efectúo una doble corrección, por parte de dos correctores independientes sobre 150 protocolos y para todas las pruebas. Los resultados del análisis de acuerdo entre correctores pueden consultarse en el capítulo 5 de este manual.

Muestra de tipificación real

La muestra de tipificación estuvo compuesta por 885 niños (frente a los 900 planeados), de edades de 2:6 a 7:7 años, repartidos en los 9 grupos de edad previstos en el diseño de la muestra. La muestra estuvo compuesta por 435 niños (49.2 %) y 450 niñas (50.8 %).

En la tabla 4.3 se muestra el tamaño de la muestra por sexo y para cada grupo de edad.

Tabla 4.3. Tamaño de la muestra de tipificación por sexo y por grupo de edad

	Tamaño de la muestra		
Grupo de edad	**Varones**	**Mujeres**	**Total**
2:6-2:11	49	49	98
3:0-3:5	47	54	101
3:6-3:11	50	48	98
4:0-4:5	51	52	103
4:6-4:11	46	50	96
5:0-5:5	50	51	101
5:6-5:11	48	46	94
6:0-6:11	46	49	95
7:0-7:7	48	51	99
Total	435	450	885

En el diseño de la muestra se puso una atención especial en que participaran niños pertenecientes a todas las comunidades autónomas de España, objetivo que se consiguió ampliamente; casi todas las comunidades españolas están representadas en la muestra de tipificación. La figura 4.1 muestra la distribución de los profesionales que recogieron los datos de la muestra de tipificación y, por tanto, las zonas de procedencia de los participantes de la muestra.

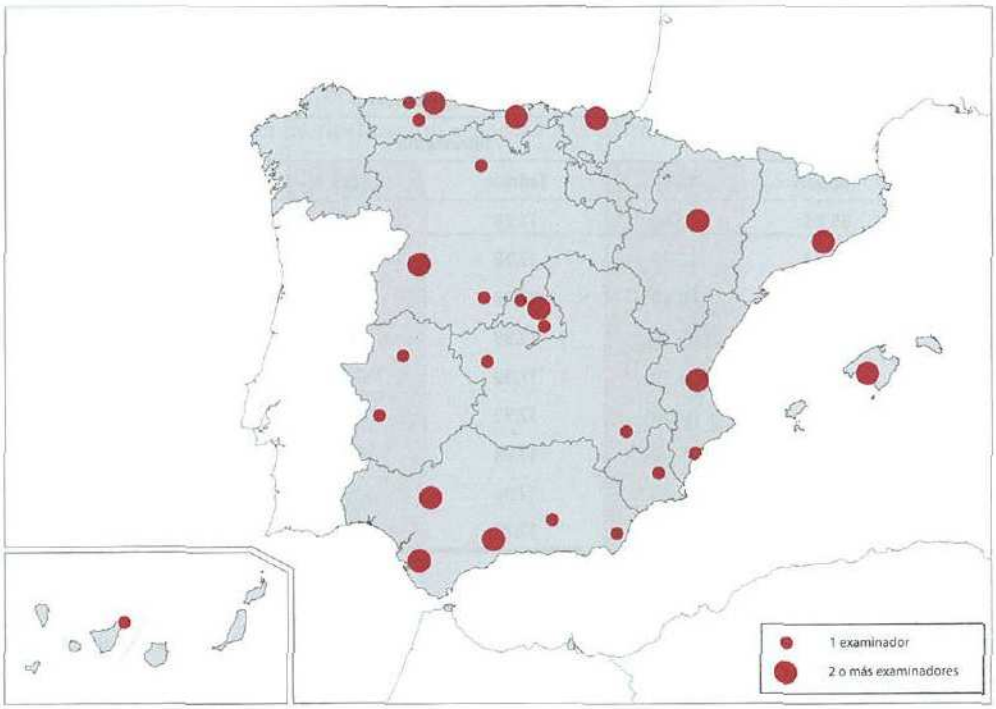


Figura 4.1. Distribución de los examinadores de la fase de tipificación

Los criterios de estratificación de las variables sexo, nivel educativo, tipo de población y zona geográfica se respetaron de forma escrupulosa por parte de los examinadores para conseguir que la representatividad de la muestra fuese muy elevada.

Representatividad de la muestra

Las tablas 4.4-4.7 muestran los porcentajes teóricos de población, basados en el censo, y los porcentajes reales conseguidos en la muestra de tipificación. Los porcentajes se presentan para las variables nivel educativo, tipo de población, zona geográfica y sexo, y en función de cada grupo de edad. Los porcentajes que se muestran a continuación están redondeados al decimal más próximo.

Tabla 4.4. Porcentajes teóricos (censo) y reales (muestra de tipificación), según edad y nivel educativo

	Nivel educativo							
	Sin estudios		Primer grado		Segundo grado		Tercer grado	
Edad	Teórico	Real	Teórico	Real	Teórico	Real	Teórico	Real
2:6-2:11	2.50	2.00	9.90	9.20	66.87	67.30	20.73	21.40
3:0-3:5	2.50	1.00	9.90	9.90	66.87	68.30	20.73	20.80
3:6-3:11	2.50	3.10	9.90	7.10	66.87	66.30	20.73	23.50
4:0-4:5	3.43	3.90	12.99	11.70	60.41	60.20	23.18	24.30
4:6-4:11	3.43	4.20	12.99	13.50	60.41	57.30	23.18	25.00
5:0-5:5	3.43	3.00	12.99	12.90	60.41	59.40	23.18	24.80
5:6-5:11	3.43	3.20	12.99	11.70	60.41	62.80	23.18	22.30
6:0-6:11	3.43	4.20	12.99	12.60	60.41	60.00	23.18	23.20
7:0-7:7	4.54	4.00	16.36	15.20	58.84	60.60	20.27	20.20

Tabla 4.5. Porcentajes teóricos (censo) y reales (muestra de tipificación), según edad y tipo de población

	Tipo de población					
	Urbana		Intermedia		Rural	
Edad	Teórico	Real	Teórico	Real	Teórico	Real
2:6-2:11	68.54	72.40	17.28	15.30	14.18	12.20
3:0-3:5	68.54	68.30	17.28	18.80	14.18	12.90
3:6-3:11	68.54	70.40	17.28	16.30	14.18	13.30
4:0-4:5	67.51	68.90	17.99	17.50	14.50	13.60
4:6-4:11	67.51	69.80	17.99	16.70	14.50	13.50
5:0-5:5	67.51	67.30	17.99	18.80	14.50	13.90
5:6-5:11	67.51	66.00	17.99	20.20	14.50	13.80
6:0-6:11	67.51	69.50	17.99	17.90	14.50	12.60
7:0-7:7	68.54	67.70	17.28	16.20	14.18	16.20

Tabla 4.6. Porcentajes teóricos (censo) y reales (muestra de tipificación), según edad y zona geográfica

	Zona geográfica							
	Norte		Sur		Este		Centro	
Edad	Teórico	Real	Teórico	Real	Teórico	Real	Teórico	Real
2:6-2:11	24.02	22.40	25.54	27.60	27.72	27.60	22.71	22.40
3:0-3:5	24.02	22.80	25.54	24.80	27.72	25.70	22.71	26.70
3:6-3:11	24.02	22.40	25.54	28.60	27.72	26.50	22.71	22.40
4:0-4:5	24.02	20.40	25.54	29.10	27.72	25.20	22.71	25.20
4:6-4:11	24.02	22.90	25.54	28.10	27.72	27.10	22.71	21.90
5:0-5:5	24.02	24.80	25.54	23.80	27.72	26.70	22.71	24.80
5:6-5:11	24.02	22.30	25.54	26.60	27.72	26.60	22.71	24.50
6:0-6:11	24.02	23.20	25.54	24.20	27.72	30.50	22.71	22.10
7:0-7:7	24.02	21.20	25.54	29.30	27.72	27.30	22.71	22.20

Tabla 4.7. Porcentajes teóricos (censo) y reales (muestra de tipificación), según edad y sexo

	Sexo			
	Varón		Mujer	
Edad	Teórico	Real	Teórico	Real
2:6-2:11	50.00	50.00	50.00	50.00
3:0-3:5	50.00	46.50	50.00	53.50
3:6-3:11	50.00	51.00	50.00	49.00
4:0-4:5	50.00	49.50	50.00	50.50
4:6-4:11	50.00	47.90	50.00	52.10
5:0-5:5	50.00	49.50	50.00	50.50
5:6-5:11	50.00	51.10	50.00	48.90
6:0-6:11	50.00	48.40	50.00	51.60
7:0-7:7	50.00	48.50	50.00	51.50

Tal y como puede observarse en las tablas, todos los porcentajes reales se ajustan de forma satisfactoria a los porcentajes teóricos basados en el censo de la población española, con ligeras

variaciones. La diferencia más grande se observa en la variable zona geográfica, para los niños del centro y para el grupo de edad de 3:0 a 3:5 años.

Los gráficos que siguen a continuación (figura 4.2) reflejan la distribución de las características demográficas de la muestra de tipificación respecto a la distribución total de la población procedente del censo.

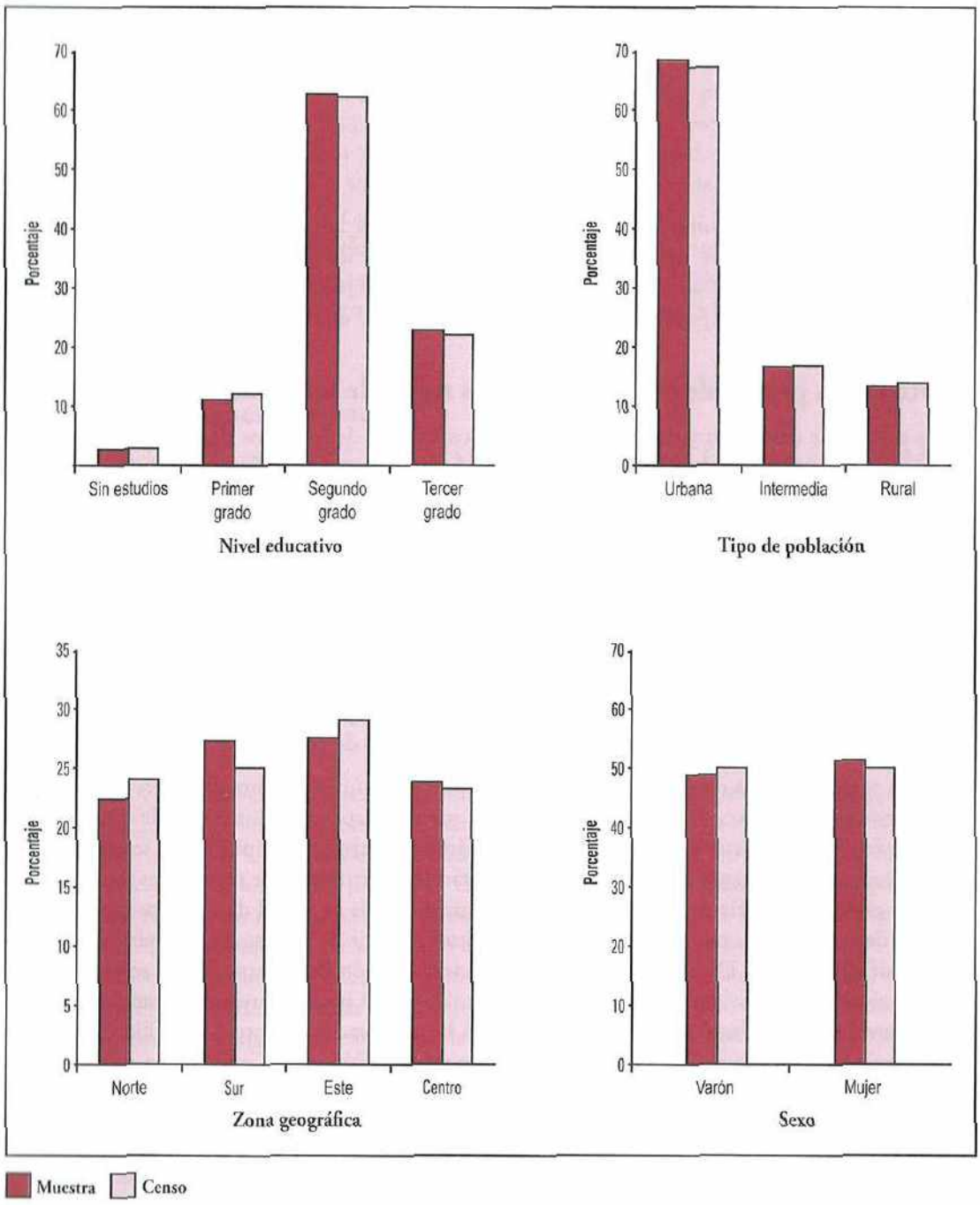


Figura 4.2. Comparación entre los datos del censo y los datos de la muestra de tipificación según las características demográficas

Como puede observarse en los gráficos, los porcentajes son extremadamente similares, especialmente para las variables nivel educativo, sexo y tipo de población. Para la zona geográfica se observaron diferencias entre lo teórico y lo real. Por ejemplo, en el caso de la región norte, se recogieron menos datos de los inicialmente planeados, mientras que en el sur participaron un porcentaje mayor de niños de los que se había planificado. Sin embargo, estas diferencias fueron mínimas.

La extrema semejanza entre los porcentajes del censo y los de la muestra de tipificación se consiguió porque cada examinador contó con una plantilla única y personalizada donde figuraban los perfiles de cada uno de los niños que debía reclutar. Aunque la elaboración de estos perfiles supuso una tarea de reclutamiento ardua, gracias al diseño personalizado y controlado en función de las características del censo y al esfuerzo de los examinadores, se consiguió una muestra de tipificación notablemente representativa de la población española.

Así, los baremos construidos a partir de los datos de la muestra de tipificación, los estudios de validez y de fiabilidad permiten una interpretación precisa y correcta de los resultados obtenidos en la WPPSI-IV.

Creación de los baremos

En este apartado se resume el procedimiento seguido para la creación de los baremos, que se inició con la preparación de los datos (establecimiento de los puntos de comienzo y reglas de terminación de las pruebas) y finalizó con la obtención de las puntuaciones típicas (de las pruebas, de procesamiento y compuestas) y las puntuaciones de edad equivalente.

Establecimiento de los puntos de comienzo y de las reglas de terminación

Los puntos de comienzo y de terminación, fijados a partir de los datos de la versión estadounidense, se han incorporado a la escala para evitar los sentimientos de frustración causados por la presencia de ítems demasiado fáciles o difíciles. Los puntos de comienzo permiten limitar el uso de las reglas de retorno y garantizar que la mayoría de los niños podrán responder correctamente a los primeros ítems de cada prueba. En la versión de la escala utilizada en la fase de tipificación los ítems de cada prueba se ordenaron siguiendo un nivel de dificultad progresivo. Los puntos de comienzo de la versión de tipificación se fijaron a partir de los datos de experimentación para reducir el número de ítems aplicados, garantizando, al mismo tiempo, que al menos el 95 % de las puntuaciones directas de las pruebas y al menos el 98 % de las puntuaciones escalares de las pruebas no cambiaban tras la aplicación de los nuevos puntos de comienzo. En la versión de tipificación y en la final se mantuvieron los mismos puntos de comienzo.

Por su parte, las reglas de terminación permiten limitar el número de errores consecutivos cometidos por el niño y garantizar, al mismo tiempo, que se ha aplicado el máximo de ítems para los que se espera un acierto. Las reglas de terminación de la versión de tipificación se establecieron generosamente para permitir que los niños tuvieran la oportunidad de realizar los ítems para los que se esperaba un acierto, aunque se limitara igualmente la cantidad de ítems presentados. Las reglas de terminación para cada prueba se ajustaron a partir de los estudios empíricos sobre los datos de tipificación. De este modo, las reglas de terminación de la versión final se establecieron como el número más bajo de respuestas consecutivas de 0 puntos cuyas puntuaciones directas mostraron una correlación de .95 o superior con las puntuaciones obtenidas utilizando la regla de terminación anterior. Por ejemplo, la regla de terminación para *Cubos* estaba establecida en tres respuestas consecutivas de 0 puntos en la versión de tipificación y se redujo a dos respuestas consecutivas de 0 puntos en la versión final de la escala. Tras esta reducción, el 99 % de las puntuaciones directas obtenidas por los niños en *Cubos* se conservó en la misma posición relativa dentro de cada grupo de edad.

Obtención de las puntuaciones escalares de las pruebas

Los datos de tipificación española se obtuvieron utilizando un método de tipificación inferencial (Wilkins, Rolfhus, Weiss y Zhu, 2005; ; Zhu y Chen, 2011). Para cada uno de los grupos de edad de la muestra de tipificación se calcularon varios parámetros de las puntuaciones (medias, desviaciones típicas y asimetrías) y se representaron gráficamente para todas las edades. Los datos de los parámetros se ajustaron a varias regresiones polinómicas, que abarcaban desde polinomios lineales hasta polinomios de cuarto grado. Las mejores curvas para cada parámetro se seleccionaron a partir de la consistencia con las expectativas teóricas subyacentes y con el patrón de las curvas de crecimiento observado en la WPPSI-IV y en otras muestras de tipificación comparables. Estas

curvas seleccionadas se usaron para obtener estimaciones de los parámetros de la población, que luego permitieron generar distribuciones teóricas para los grupos de edad normativos, ofreciendo percentiles para cada puntuación directa. Mediante el método del percentil medio del intervalo, estos percentiles se convirtieron en puntuaciones escalares con una media de 10, una desviación típica de 3 y un rango de 1-19. Seguidamente, se examinó la progresión de las puntuaciones escalares dentro de cada grupo de edad y de un grupo de edad a otro, y las pequeñas irregularidades se corrigieron mediante suavizado.

La tabla A.1 del *Manual de aplicación y corrección* ofrece las puntuaciones escalares correspondientes a las puntuaciones directas de cada prueba, así como dos puntuaciones de procesamiento, para cada grupo de edad. Las puntuaciones escalares equivalentes se presentan en intervalos de 3 meses para las edades comprendidas entre los 2 años, 6 meses y 0 días y los 5 años, 11 meses y 30 días; así como en intervalos de 4 meses para las edades comprendidas entre los 6 años, 0 meses y 0 días y los 7 años, 7 meses y 30 días. Para las pruebas *Información, Dibujos, Nombres, Cubos, Rompecabezas, Reconocimiento* y *Localización* se establecieron baremos para ambas franjas de edad. Para las pruebas *Semejanzas, Vocabulario, Comprensión, Matrices, Conceptos, Búsqueda de animales, Cancelación* y *Clave de figuras* se establecieron baremos solo para la franja de más edad. El rango de las puntuaciones escalares obtenidas es de 1 a 19, con una desviación típica de 3 a cada lado de la media.

Obtención de las puntuaciones compuestas

Las puntuaciones compuestas (ICV, IVE, IMT, CIT, IAV, INV e ICG, para el grupo de edad de 2:6 a 3:11 años; e ICV, IVE, IRF, IMT, IVP, CIT, IAV, INV, ICG e ICC para el grupo de edad de 4:0 a 7:7 años) se basan en la suma correspondiente de las puntuaciones escalares ajustadas a la edad. Para elaborar las tablas de puntuaciones compuestas, se calculó la suma de las puntuaciones escalares de cada puntuación compuesta para cada niño de la muestra de tipificación, utilizando las puntuaciones escalares de las pruebas principales que contribuyen a esa puntuación compuesta.

Las puntuaciones compuestas que se obtienen con distintas combinaciones de pruebas, en función de la franja de edad, se tipificaron por separado. En las tablas 2.5 y 2.6 del *Manual de aplicación y corrección* se presentan las pruebas que contribuyen a las puntuaciones compuestas para cada franja de edad.

Las medias y las desviaciones típicas de la suma de las puntuaciones escalares de cada puntuación compuesta figuran, para cada grupo de edad tipificado, en las tablas A.1-A.9 de este manual. Estos datos son muy similares de una edad a otra, y el análisis de varianza no reveló diferencias significativas de las sumas promedio de las puntuaciones escalares para cada puntuación compuesta en función de la edad. Asimismo, los resultados del test de Bartlett para la homogeneidad de varianzas aplicado a cada grupo de edad revelaron que las varianzas de las sumas de las puntuaciones escalares no diferían significativamente de una edad a otra. Además, el examen de las distribuciones de las sumas de las puntuaciones escalares mostró que estas sumas estaban normalmente distribuidas. En consecuencia, los grupos de edad de cada franja de edad fueron combinados para elaborar las tablas de puntuaciones compuestas equivalentes.

Para cada escala se utilizó la distribución de la suma de las puntuaciones escalares para obtener los percentiles en el medio del intervalo correspondiente, que se convirtieron en puntuaciones típicas con una media 100 y una desviación típica 15, y un rango de 45-155 para las escalas Comprensión verbal, Visoespacial, Razonamiento fluido, Memoria de trabajo, Velocidad de procesamiento y Adquisición de vocabulario; y un rango de 40-160 para la Escala total y las escalas No verbal, Capacidad general y Competencia cognitiva. Las distribuciones de las puntuaciones típicas resultantes se suavizaron visualmente para eliminar posibles irregularidades y garantizar que la distribución era aproximadamente normal, tratando en todo momento de mantener las medias y desviaciones típicas de las escalas próximas a 100 y 15, respectivamente (Allen y Yen, 2002; Cureton y Tukey, 1951). El rango de puntuaciones de los índices es de 45-155 para el ICV, IVE, IRF, IMT, IVP e IAV; y de 40-160 para el CIT, INV, ICG e ICC.

En las tablas A.2-A.11 del *Manual de aplicación y corrección* se presenta la puntuación compuesta correspondiente a las sumas de las puntuaciones escalares, por franja de edad, para la Escala total y las escalas Comprensión verbal, Visoespacial y Memoria de trabajo; y, para la franja de más

edad, la puntuación compuesta correspondiente a las sumas de las puntuaciones escalares para las escalas Razonamiento fluido y Velocidad de procesamiento. En la tabla B.1 de este manual figura la puntuación compuesta correspondiente a las sumas de las puntuaciones escalares de la escala Adquisición de Vocabulario; y en las tablas B.2 y B.3, la puntuación compuesta correspondiente a las sumas de las puntuaciones escalares, por franja de edad, de la escala No verbal. Y, finalmente, en las tablas C.1 y C.2 de este manual se presenta la puntuación compuesta correspondiente a las sumas de las puntuaciones escalares, por franja de edad, de la escala Capacidad general; y en la tabla C.3, la puntuación compuesta correspondiente a las sumas de las puntuaciones escalares de la escala Competencia cognitiva para el grupo de más edad.

Cálculo de las puntuaciones de edad equivalente

La edad equivalente de una puntuación directa indica la edad en la que dicha puntuación es igual a la media. A fin de obtener las edades equivalentes de una prueba, se determinó la puntuación directa correspondiente a una puntuación escalar de 10 para cada uno de los 19 grupos de edad tipificados, y dicha puntuación directa se fijó como punto medio del rango de edad de cada uno de los grupos (p. ej., 2:7 para el grupo de edad de 2:6 a 2:8 años). Si la misma puntuación directa se correspondía con dos grupos de edad sucesivos (p. ej., 2:6-2:8 y 2:9-2:11), se asignaba al rango del grupo de menor edad (p. ej., 2:6–2:8). Si la misma puntuación directa se correspondía con tres grupos de edad sucesivos (p. ej., 2:6–2:8, 2:9-2:11 y 3:0-3:2), se asignaba al rango del grupo de edad intermedia (p. ej., 2:9-2:11). Si el mismo rango de puntuaciones directas se correspondía con una puntuación escalar de 10 en grupos adyacentes (p. ej., puntuaciones directas de 18 y 19 que se corresponden a una puntuación escalar de 10 en grupos de edad adyacentes), se dividían los valores del rango a partes iguales entre los grupos de edad adyacentes.

En la tabla A.13 del *Manual de aplicación y corrección* se presentan las edades equivalentes a las puntuaciones directas de cada prueba. Ha de tenerse en cuenta que las edades equivalentes a menudo se malinterpretan y tienen limitaciones psicométricas; por ello, deben utilizarse con cautela. Las edades equivalentes no ofrecen información comparativa de un mismo grupo de edad; esta información se obtiene mediante las puntuaciones típicas y los percentiles. Las edades equivalentes son medias de los distintos grupos de edad y no proporcionan información sobre la distribución de las puntuaciones en un determinado grupo.

Una edad equivalente de un niño puede sugerir la presencia de un déficit cuando, de hecho, el niño se sitúa en el rango medio de funcionamiento para su edad. En el capítulo 6 de este manual se ofrece más información sobre las limitaciones y el uso adecuado de las edades equivalentes.

Fiabilidad

Las propiedades psicométricas de la WPPSI-IV determinan el nivel de confianza que los profesionales pueden depositar en la exactitud y precisión de las puntuaciones obtenidas. Para realizar un adecuado análisis e interpretación de los resultados logrados tras la aplicación de la escala, es fundamental que el profesional conozca la fiabilidad y estabilidad, los errores típicos de medida, los intervalos de confianza, la significación estadística de las diferencias entre puntuaciones (valores críticos) y la frecuencia con la que aparecen esas diferencias (tasas base).

Las limitaciones que presenta la prueba de significación de la hipótesis nula (p) y la toma de decisiones basada únicamente en este valor han sido extensamente documentadas (Balluerka, Gómez e Hidalgo, 2005; Cohen, 1990, 1994; Dunst y Hamby, 2012; Hubbard y Lindsay, 2008; Hunter, 1997; Kirk, 1996, 2001; Thompson, 1998). Esto ha llevado a que, además de informar de las puntuaciones obtenidas y de la significación estadística, sea necesario recurrir también a los intervalos de confianza y al tamaño del efecto (American Educational Research Association [AERA], 2006). Por ello, en este manual se presentan los intervalos de confianza y el tamaño del efecto, así como las pruebas de significación estadística y el valor de p, a fin de demostrar la fiabilidad y la validez de la escala y para que el profesional entienda mejor el significado de las puntuaciones obtenidas.

En la presentación de los resultados psicométricos se han seguido una serie de criterios. Por una parte, todos los análisis usan los valores tradicionales para el nivel de significación (alfa = .05). Por otra parte, en este manual, el término *diferencia típica* se refiere a la d de Cohen. Los valores del tamaño del efecto se interpretan siguiendo las sugerencias de Cohen (1988, 1992): un tamaño del efecto de .20 se considera pequeño; de .50, moderado; y de .80 grande. Aunque estas referencias no describen por completo todos los aspectos de la interpretación del tamaño del efecto, en atención a la simplicidad y a la consistencia con anteriores versiones de las escalas de inteligencia de Wechsler, los valores de d comprendidos entre .20 y .49 se consideran pequeños, los comprendidos entre .50 y .79 se consideran moderados, y los valores de .80 o superiores se consideran grandes. Véase Cohen (1988) para más información sobre la interpretación de los tamaños del efecto.

Si bien en este manual se siguen los criterios habituales respecto a la significación estadística y a la interpretación del tamaño del efecto, siempre deben tenerse en cuenta los valores observados. Por ejemplo, aunque un valor de p = .07 no resulte estadísticamente significativo, se debe valorar de distinta manera si el tamaño del efecto que tiene asociado es de .60, que si es de .10. Al contrario, si un valor de p = .04 resulta estadísticamente significativo, pero tiene asociado un tamaño del efecto de .10, el efecto es demasiado pequeño para ser relevante en la práctica.

Las sugerencias de Cohen han de entenderse como directrices generales, y, en función de la situación, la importancia asociada a un tamaño del efecto determinado podrá ser muy distinta a la sugerida por Cohen. En ciertas situaciones clínicas, un tamaño del efecto pequeño podría llegar a representar un importante hallazgo. Se recomienda al profesional que evalúe los detalles de un determinado resultado cuando interprete la significación y el tamaño del efecto.

Fiabilidad y errores de medida

La fiabilidad de un test se refiere a la exactitud, consistencia y estabilidad de las puntuaciones en diferentes situaciones (Anastasi y Urbina, 1998; Sattler, 2008a). La teoría clásica de los tests considera que la puntuación obtenida en un test es una aproximación a la *puntuación verdadera* hipotética de un niño, es decir, a la puntuación que el niño obtendría si el test fuera totalmente fiable. La diferencia entre la puntuación verdadera hipotética y la puntuación obtenida en la prueba es lo que se denomina *error de medida*. Un test fiable tiene errores de medida relativamente pequeños y proporciona resultados consistentes tanto si se aplica una sola vez como si se aplica en diferentes ocasiones. La fiabilidad de un test siempre debería tenerse en cuenta en la interpretación de las puntuaciones obtenidas por el niño y de las diferencias encontradas entre las puntuaciones obtenidas por ese mismo niño en distintas ocasiones (p. ej., test-retest).

Consistencia interna

Los coeficientes de fiabilidad se calcularon utilizando el método de las dos mitades. El coeficiente de fiabilidad basado en el método de las dos mitades es la correlación entre las puntuaciones de las dos mitades corregida mediante la fórmula de Spearman-Brown para cada prueba (Crocker y Algina, 1986; Li, Rosenthal y Rubin, 1996).

Dado que *Búsqueda de animales*, *Cancelación* y *Clave de figuras* son pruebas de Velocidad de procesamiento, el coeficiente de las dos mitades no se considera un método adecuado de valoración de la fiabilidad. Por ello, para valorar la fiabilidad de estas pruebas, se usaron los coeficientes de estabilidad test-retest. Estos coeficientes de estabilidad se basaban en las puntuaciones obtenidas en el análisis test-retest, como se describe más adelante en este capítulo. El coeficiente de estabilidad es la correlación entre las puntuaciones de la primera y segunda aplicación, corregido de acuerdo con la variabilidad de la muestra de tipificación (Allen y Yen, 2002; Magnusson, 1967).

Los coeficientes de fiabilidad de consistencia interna para las puntuaciones compuestas se calcularon con la fórmula recomendada por Guilford (1954) y por Nunnally y Bernstein (1994). En la tabla 5.1 figuran los coeficientes de fiabilidad de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas, para toda la muestra y para cada grupo y franja de edad. Los coeficientes de fiabilidad promedio se calcularon utilizando la trasformación *z* de Fisher (Silver y Dunlap, 1987; Strube, 1988).

Tabla 5.1. Coeficientes de fiabilidad de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas (datos españoles)

	Edad											
	2:6–3:11				4:0–7:7							
Prueba/ punt. de procesamiento/ punt. compuesta	2:6–2:11	3:0–3:5	3:6–3:11	Promedio r_{xx}[a]	4:0–4:5	4:6–4:11	5:0–5:5	5:6–5:11	6:0–6:11	7:0–7:7	Promedio r_{xx}[a]	Total promedio r_{xx}[a]
Información	.83	.87	.78	.83	.76	.83	.75	.72	.70	.86	.78	.80
Semejanzas	–	–	–	–	.88	.86	.78	.83	.80	.83	.83	.83
Vocabulario	–	–	–	–	.84	.87	.76	.81	.78	.82	.82	.82
Comprensión	–	–	–	–	.86	.83	.78	.76	.62	.78	.78	.78
Dibujos	.83	.88	.88	.87	.90	.78	.80	.84	.70	.79	.81	.83
Nombres	.81	.88	.88	.86	.85	.84	.80	.85	.63	.64	.78	.81
Cubos	.85	.86	.82	.84	.83	.72	.79	.78	.77	.82	.79	.81
Rompecabezas	.84	.88	.87	.86	.87	.80	.77	.75	.64	.62	.76	.80
Matrices	–	–	–	–	.85	.90	.82	.87	.83	.83	.85	.85
Conceptos	–	–	–	–	.90	.86	.83	.87	.81	.73	.84	.84
Reconocimiento	.86	.83	.88	.86	.84	.84	.85	.90	.84	.90	.86	.86
Localización	.89	.80	.85	.85	.82	.76	.79	.82	.65	.77	.77	.80
Búsqueda de animales	–	–	–	–	.92	.92	.92	.78	.78	.78	.87	.87
Cancelación	–	–	–	–	.77	.77	.77	.77	.77	.77	.77	.77
Clave de figuras	–	–	–	–	.87	.87	.87	.68	.68	.68	.79	.79
Cancelación aleat.	–	–	–	–	.63	.63	.63	.79	.79	.79	.72	.72
Cancelación estruct.	–	–	–	–	.81	.81	.81	.61	.61	.61	.72	.72
ICV	.88	.91	.88	.89	.88	.89	.85	.85	.80	.88	.86	.87
IVE	.88	.90	.89	.89	.89	.84	.85	.84	.79	.75	.83	.85
IRF	–	–	–	–	.90	.91	.88	.92	.86	.85	.89	.89
IMT	.92	.86	.91	.90	.88	.82	.85	.88	.80	.88	.85	.87
IVP	–	–	–	–	.90	.88	.89	.84	.82	.85	.87	.87
CIT	.93	.93	.93	.93	.94	.93	.94	.93	.88	.93	.93	.93
IAV	.85	.91	.92	.90	.91	.87	.87	.88	.77	.81	.86	.87
INV	.93	.91	.93	.92	.94	.92	.94	.93	.88	.92	.92	.92
ICG	.91	.93	.91	.92	.92	.91	.91	.91	.87	.92	.91	.91
ICC	–	–	–	–	.92	.87	.91	.90	.86	.90	.90	.90

[a] Los coeficientes de fiabilidad promedio se calcularon con la transformación *z* de Fisher.

Como muestran los datos de la tabla 5.1, los coeficientes de fiabilidad promedio de las pruebas, para la muestra de tipificación española, se sitúan entre adecuados (*Cancelación, Clave de figuras* y *Comprensión*), buenos (*Información, Semejanzas, Vocabulario, Dibujos, Nombres, Cubos, Rompecabezas, Conceptos,* y *Localización*) y excelentes (*Matrices, Reconocimiento* y *Búsqueda de animales*). En general, la fiabilidad de las pruebas de la WPPSI-IV conservan o mejoran la fiabilidad de las pruebas de la WPPSI-III. Aunque la WPSSI-III no incluía pruebas de Memoria de trabajo con las que comparar las de la WPPSI-IV, los coeficientes de fiabilidad de *Reconocimiento* y *Localización* reflejan una calidad similar a la de los coeficientes del resto de las pruebas de la WPPSI-III. Los coeficientes de fiabilidad de las pruebas nuevas de Velocidad de procesamiento se basan en las fiabilidades test-retest, que suelen ser inferiores a las fiabilidades de las dos mitades porque las primeras presentan fuentes de error adicionales, relacionadas con posibles cambios en la conducta del niño entre las dos sesiones de aplicación. No obstante, el coeficiente de fiabilidad de *Búsqueda de animales* es superior al de la prueba de la WPPSI-III a la que reemplazó, *Búsqueda de símbolos.* Los coeficientes de fiabilidad de *Cancelación* y *Clave de figuras* son ligeramente inferiores, pero aceptables, a los de la prueba *Claves* de la WPPSI-III. En conjunto, los coeficientes de fiabilidad de las pruebas de la WPPSI-IV son, por lo general, buenos o excelentes, de acuerdo con los estándares propuestos por Prieto y Muñiz (2000).

Los coeficientes de fiabilidad promedio de las puntuaciones compuestas de los índices primarios de la WPPSI-IV oscilan entre .85 y .93 y generalmente son más altos que los de cada una de las pruebas que los conforman. Esta diferencia, ya esperada, se explica porque cada prueba representa solo una pequeña parte de todo el funcionamiento intelectual de un niño, mientras que las puntuaciones compuestas (p. ej., índices primarios) resumen el rendimiento de un niño en un ámbito más amplio de aptitudes. Los altos coeficientes de fiabilidad promedio de las puntuaciones compuestas de la WPPSI-IV eran los que se esperaban. Cabe destacar que el coeficiente de fiabilidad del índice nuevo, el IMT, es bueno. El coeficiente de fiabilidad promedio del IVP, es igual a Velocidad de procesamiento de la WPPSI-III. Los coeficientes de fiabilidad del ICV y del IVE son ligeramente inferiores a los de los índices correspondientes de la WPPSI-III (CI verbal y CI manipulativo, respectivamente), y ello es debido a que las pruebas que contribuyen a cada uno de estos dos índices primarios de la WPPSI-IV son solo dos para los niños de 4:0 a 7:7 años, mientras que las pruebas que contribuyen a las puntuaciones compuestas de la WPPSI-III son tres. En cambio, el IRF, a pesar de estar compuesto solo por dos pruebas, mantiene el mismo nivel de fiabilidad que el CIM de la WPPSI-III, dado que la fiabilidad estimada para *Matrices* y *Conceptos* es superior a la de las pruebas *Cubos* y *Rompecabezas*. Un patrón similar de fiabilidad más baja estimada para las pruebas que requieren aptitud motora se observó en revisiones anteriores (Wechsler, 2002, 2003, 2008). Los coeficientes de fiabilidad promedio de los índices primarios son similares (ICV e IVE) o iguales (IRF e IVP) a las puntuaciones compuestas correspondientes de la WPPSI-III.

El coeficiente de fiabilidad del CIT es excelente y tiene el mismo valor que el de la WPPSI-III. Los coeficientes de fiabilidad de las puntuaciones de los índices secundarios también son excelentes. El coeficiente de fiabilidad del IAV tiene el mismo valor que el de la puntuación compuesta correspondiente de la WPPSI-III (Lenguaje general). El INV, el ICG y el ICC muestran una fiabilidad similar a la del CIT.

Los resultados de fiabilidad en poblaciones especiales respaldan la generalizabilidad de la escala. Para la WPPSI-IV se obtuvieron resultados de la consistencia interna de las pruebas en una muestra estadounidense, para los siguientes grupos especiales: alta capacidad intelectual, discapacidad intelectual leve, discapacidad intelectual moderada, retraso del desarrollo cognitivo, factores de riesgo para el desarrollo, problemas previos a la lectoescritura, TDAH, comportamiento perturbador, trastorno del lenguaje expresivo, trastorno mixto del lenguaje receptivo-expresivo, trastorno autista y trastorno de Asperger. En el capítulo 6 de este manual se ofrece la información demográfica detallada para estos grupos especiales, y en el anexo D consta una completa descripción de los criterios de inclusión para cada grupo.

En la tabla 5.2 se presentan los coeficientes de fiabilidad de consistencia interna de las pruebas para los grupos especiales. Los coeficientes de fiabilidad se calcularon utilizando la misma metodología descrita para los análisis que se presentan en la tabla 5.1. Para las pruebas de Velocidad de procesamiento (*Búsqueda de animales, Cancelación* y *Clave de figuras*) no se presentan los coeficientes de fiabilidad de los grupos especiales puesto que estos grupos no participaron en el estudio test-retest.

Tabla 5.2. Coeficientes de fiabilidad de las pruebas para los grupos especiales (datos estadounidenses)

	Alta capacidad intelectual		Discapacidad intelectual leve		Discapacidad intelectual moderada		Retraso del desarrollo cognitivo		Factores de riesgo para el desarrollo		Problemas previos a la lectoescritura		TDAH		Conducta disruptiva		Trastorno del lenguaje expresivo		Trastorno del lenguaje mixto receptivo-expresivo		Trastorno autista		Trastorno de Asperger		Promedio
Prueba	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	r_{xx}[a]
Información	56	.83	39	.95	33	.93	71	.97	42	.98	37	.83	53	.96	26	.81	23	.87	39	.95	38	.98	38	.87	.94
Semejanzas	55	.90	31	.97	27	.95	36	.96	33	.97	37	.94	50	.95	25	.95	25	.96	42	.96	34	.95	37	.95	.95
Vocabulario	55	.92	31	.89	26	.87	36	.91	33	.97	37	.89	50	.89	26	.90	25	.89	39	.91	34	.96	37	.94	.92
Comprensión	56	.87	31	.96	27	.81	36	.94	33	.98	37	.91	50	.93	26	.93	25	.94	40	.93	33	.96	37	.95	.94
Dibujos	56	.90	39	.97	34	.93	72	.95	42	.98	37	.85	53	.94	26	.80	25	.92	42	.95	38	.93	38	.95	.94
Nombres	56	.88	39	.95	33	.97	72	.96	41	.97	37	.86	53	.94	26	.95	25	.91	40	.93	38	.97	38	.95	.94
Cubos	56	.88	38	.96	33	.91	70	.96	41	.97	36	.76	53	.89	26	.85	25	.93	42	.93	37	.92	37	.91	.92
Rompecabezas	56	.82	39	.90	34	.89	72	.96	42	.96	37	.81	53	.94	26	.84	25	.82	42	.96	38	.94	38	.88	.91
Matrices	56	.87	31	.94	27	.89	36	.90	33	.98	37	.94	50	.92	26	.93	25	.92	42	.95	34	.97	37	.94	.94
Conceptos	55	.88	31	.97	27	.94	36	.95	33	.97	37	.94	49	.95	26	.94	25	.94	42	.95	34	.96	37	.97	.95
Reconocimiento	55	.90	39	.95	34	.90	72	.95	42	.98	37	.91	53	.93	26	.97	25	.96	42	.90	37	.97	38	.95	.95
Localización	56	.81	39	.94	34	.85	72	.95	41	.96	37	.81	53	.95	26	.88	25	.91	41	.94	38	.95	38	.92	.92

Nota. Sobre *Búsqueda de animales, Cancelación* y *Clave de figuras* no se ofrece información porque no se recogieron datos del test-retest para los grupos especiales.

[a] Los coeficientes de fiabilidad promedio se calcularon con la transformación *z* de Fisher.

Errores típicos de medida e intervalos de confianza

El error típico de medida (*ETM*) proporciona una estimación de la cantidad de error que contiene una puntuación obtenida por un niño en un test. Dado que el *ETM* es inversamente proporcional a la fiabilidad, cuanto mayor es la fiabilidad del instrumento, menor es el *ETM*, lo que aumenta la precisión de los resultados obtenidos en el test. El *ETM* es la desviación típica de la distribución del error de medida y normalmente se expresa en términos de unidades de puntuación típica. El *ETM* se calcula a partir de la siguiente fórmula:

$$ETM = Dt\sqrt{1 - r_{xx}}$$

donde *Dt* es la unidad de desviación típica del test en la escala de medida empleada y r_{xx} es el coeficiente de fiabilidad que tiene esa medida. No se deben comparar los *ETM* de las puntuaciones de las pruebas con los de las puntuaciones compuestas porque se basan en diferentes desviaciones típicas. Dado que la desviación típica de las puntuaciones de las pruebas es 3 y la de las puntuaciones compuestas es 15, los *ETM* de las puntuaciones de las pruebas suelen ser menores que los de las puntuaciones compuestas. Sin embargo, en realidad las puntuaciones compuestas suelen ser más fiables que las puntuaciones de las pruebas. En la tabla 5.3. figuran los *ETM* de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas, para toda la muestra y para cada grupo y franja de edad.

El *ETM* permite estimar los intervalos de confianza en los que se espera encontrar la verdadera puntuación del niño. El intervalo de confianza es otro modo de expresar la precisión de las puntuaciones de los tests. La puntuación obtenida en un test no es la puntuación verdadera de ese niño, por lo que los intervalos de confianza nos indican que el error de medida es inherente a la medición, y que la puntuación observada de un test es solo una estimación de la puntuación verdadera, es decir, de la capacidad real del niño. Los profesionales deben utilizar este intervalo para presentar la puntuación del niño dentro de un rango en el que probablemente se sitúa la verdadera puntuación del niño.

Para calcular los intervalos de confianza, pueden utilizarse distintos métodos, y cada uno de ellos produce valores ligeramente diferentes. En todas las fórmulas para calcular los intervalos de confianza se utiliza el *ETM*. En algunas ocasiones, los profesionales prefieren calcular los intervalos de confianza de las puntuaciones compuestas utilizando solo la puntuación observada y el *ETM*. Los intervalos de confianza basados en la puntuación observada y el *ETM* se calculan mediante la siguiente fórmula:

$$p\% \text{ intervalo de confianza} = \text{puntuación observada} \pm z_p(ETM)$$

donde p es el nivel de confianza, como 90 o 95 %, y donde z_p es el valor z asociado al nivel de confianza, que puede obtenerse en una tabla de áreas bajo la curva normal.

Para calcular los intervalos de confianza centrados en las puntuaciones observadas, el profesional, en función de su preferencia, puede usar el ETM para el grupo de edad o para la franja de edad del niño. En los ejemplos de este apartado se han utilizado los *ETM* basados en el grupo de edad del niño. Por ejemplo, si un niño de 6 años obtiene un CI total de 108, el profesional puede estar seguro en un 95 % de que el verdadero CI total del niño se sitúa en el rango de 102-114 (porque el intervalo de confianza del 95 % es 108 ± 1.96 *ETM*, donde el *ETM* es 3.02), y en un 90 % de que el CI total del sujeto se encuentra en el rango de 103-113 (108 ± 1.65 *ETM*).

Tabla 5.3. Errores típicos de medida de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas (datos españoles)

Prueba/ *punt. de procesamiento*/ punt. compuesta	Edad 2:6–3:11				Edad 4:0–7:7							Total promedio *ETM*[a]
	2:6–2:11	3:0–3:5	3:6–3:11	Promedio *ETM*[a]	4:0–4:5	4:6–4:11	5:0–5:5	5:6–5:11	6:0–6:11	7:0–7:7	Promedio *ETM*[a]	
Información	1.24	1.08	1.41	1.25	1.47	1.24	1.50	1.59	1.64	1.12	1.44	1.38
Semejanzas	–	–	–	–	1.04	1.12	1.41	1.24	1.34	1.24	1.24	1.24
Vocabulario	–	–	–	–	1.20	1.08	1.47	1.31	1.41	1.27	1.30	1.30
Comprensión	–	–	–	–	1.12	1.24	1.41	1.47	1.85	1.41	1.43	1.43
Dibujos	1.24	1.04	1.04	1.11	0.95	1.41	1.34	1.20	1.64	1.37	1.33	1.26
Nombres	1.31	1.04	1.04	1.14	1.16	1.20	1.34	1.16	1.82	1.80	1.44	1.35
Cubos	1.16	1.12	1.27	1.19	1.24	1.59	1.37	1.41	1.44	1.27	1.39	1.33
Rompecabezas	1.20	1.04	1.08	1.11	1.08	1.34	1.44	1.50	1.80	1.85	1.52	1.40
Matrices	–	–	–	–	1.16	0.95	1.27	1.08	1.24	1.24	1.16	1.16
Conceptos	–	–	–	–	0.95	1.12	1.24	1.08	1.31	1.56	1.23	1.23
Reconocimiento	1.12	1.24	1.04	1.14	1.20	1.20	1.16	0.95	1.20	0.95	1.12	1.12
Localización	0.99	1.34	1.16	1.17	1.27	1.47	1.37	1.27	1.77	1.44	1.44	1.36
Búsqueda de animales	–	–	–	–	0.85	0.85	0.85	1.41	1.41	1.41	1.16	1.16
Cancelación	–	–	–	–	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44
Clave de figuras	–	–	–	–	1.08	1.08	1.08	1.70	1.70	1.70	1.42	1.42
Cancelación aleat.	–	–	–	–	1.82	1.82	1.82	1.37	1.37	1.37	1.61	1.61
Cancelación estruct.	–	–	–	–	1.31	1.31	1.31	1.87	1.87	1.87	1.61	1.61
ICV	5.20	4.50	5.20	4.98	5.20	4.97	5.81	5.81	6.71	5.20	5.65	5.43
IVE	5.20	4.74	4.97	4.97	4.97	6.00	5.81	6.00	6.87	7.50	6.24	5.85
IRF	–	–	–	–	4.74	4.50	5.20	4.24	5.61	5.81	5.05	5.05
IMT	4.24	5.61	4.50	4.82	5.20	6.36	5.81	5.20	6.71	5.20	5.78	5.48
IVP	–	–	–	–	4.74	5.20	4.97	6.00	6.36	5.81	5.54	5.54
CIT	3.97	3.97	3.97	3.97	3.67	3.97	3.67	3.97	5.20	3.97	4.11	4.06
IAV	5.81	4.50	4.24	4.90	4.50	5.41	5.41	5.20	7.19	6.54	5.78	5.50
INV	3.97	4.50	3.97	4.15	3.67	4.24	3.67	3.97	5.20	4.24	4.20	4.18
ICG	4.50	3.97	4.50	4.33	4.24	4.50	4.50	4.50	5.41	4.24	4.58	4.50
ICC	–	–	–	–	4.24	5.41	4.50	4.74	5.61	4.74	4.90	4.90

Nota. Los *ETM* se presentan en puntuaciones escalares para las pruebas y en puntuaciones compuestas para los índices y el CI total. Para calcular estos *ETM*, se utilizaron los coeficientes de fiabilidad que figuran en la tabla 5.1 y las desviaciones típicas (a saber, 3 para las pruebas y 15 para las *puntuaciones compuestas*).

[a] Los promedios de los *ETM* se calcularon promediando la suma de los *ETM* elevados al cuadrado en cada grupo de edad, y obteniendo la raíz cuadrada del resultado.

Los intervalos de confianza de los rangos de puntuaciones compuestas que figuran en las tablas A.2-A.11 del *Manual de aplicación y corrección* y en la tablas B.1-B.3 y C.1-C.3 de este manual se obtuvieron mediante las puntuaciones verdaderas estimadas (*PVE*) y el error típico de estimación (*ETE*), de acuerdo con el método propuesto por Dudek (1979) y Glutting, McDermott y Stanley (1987). La puntuación verdadera estimada se obtiene mediante la fórmula:

$$\text{PVE} = 100 + r_{xx}(X - 100)$$

donde X es la puntuación compuesta observada, y *r*xx es el coeficiente de fiabilidad de la puntuación compuesta. El *ETE* se deriva con la fórmula propuesta por Stanley (1971):

$$ETE = Dt\left(r_{xx}\right)\sqrt{1 - r_{xx}}$$

donde *Dt* es la desviación típica teórica de la puntuación compuesta, y *r*xx es el coeficiente de fiabilidad de la puntuación compuesta.

Este método centra el intervalo de confianza en la puntuación verdadera estimada, en lugar de hacerlo en la puntuación observada, lo que da lugar a un intervalo asimétrico respecto a la puntuación observada. Esta asimetría se produce porque la puntuación verdadera estimada esta más cerca de la media de la escala que la puntuación observada. Un intervalo de confianza basado en la puntuación verdadera estimada y en el *ETE* es una corrección de la regresión de la puntuación verdadera respecto a la media. Por ejemplo, si un niño de 7 años obtiene un CI total de 112, su puntuación verdadera estimada será de 111.2, el intervalo de confianza del 95 % del CI total verdadero del niño se situará en el rango de 104-118 (porque el intervalo de confianza del 95 % es 111.2 ± 1.96 *ETE*, donde el *ETE* es 3.69), y el intervalo de confianza del 90% del CI total verdadero del niño se encontrará en el rango de 105-117 (111.2 ± 1.65 *ETE*).

Además, los intervalos de confianza se calcularon utilizando los promedios del coeficiente de fiabilidad para cada franja de edad, en vez de para cada grupo de edad, porque los coeficientes de fiabilidad de las puntuaciones compuestas y los *ETE* son muy similares entre los grupos de edad. Del mismo modo, dado que los *ETE* de los grupos de edad y de las franjas de edad no son muy diferentes, los intervalos de confianza centrados en la puntuación obtenida y calculados con uno de los dos valores serían iguales o muy similares. Asimismo, dado que los coeficientes de fiabilidad de las puntuaciones compuestas son relativamente elevados, los intervalos de confianza calculados con el *ETM* centrado en la puntuación obtenida y los intervalos de confianza calculados con el *ETE* centrado en la puntuación verdadera estimada serán iguales o similares. En algunos casos, los profesionales pueden querer calcular los intervalos de confianza centrados en la puntuación observada, usar niveles de confianza distintos a los indicados en este manual o utilizar las fiabilidades del grupo de edad del niño. Los intervalos de confianza calculados usando el *ETM* o el *ETE*, o los valores del grupo de edad o de la franja de edad, se interpretan del mismo modo.

Estabilidad temporal: test-retest

La prueba de estabilidad test-retest para las pruebas, las puntuaciones de procesamiento y las puntuaciones compuestas se obtuvo a partir de una muestra española de 72 niños, de 2:6 a 7:7 años. A los participantes se les aplicó la WPPSI-IV en dos ocasiones, con un intervalo temporal de entre 7 y 42 días y un intervalo medio de 23 días. La tabla 5.4. presenta los datos demográficos del estudio test-retest y del estudio de fiabilidad por acuerdo entre correctores, que se describe más adelante. En ella figuran la edad media y los porcentajes de representación de la muestra por sexo, nivel educativo de los padres, zona geográfica y tipo de población.

Tabla 5.4. Datos demográficos del estudio de estabilidad test-retest y del estudio de acuerdo entre correctores (datos españoles)

	Estabilidad test-retest	Acuerdo entre correctores
N	72	150
Edad		
Media	4.6	5
Dt	1.4	1.5
Sexo		
Mujer	51.4	51.3
Varón	48.6	48.7
Nivel educativo de los padres		
Sin estudios	4.2	1.3
Primer grado	11.1	7.3
Segundo grado	63.9	66.7
Tercer grado	20.8	24.7
Zona geográfica		
Norte	9.7	28
Sur	33.3	26.7
Este	16.7	20
Centro	40.3	25.3
Tipo de población		
Urbana	70.8	67.3
Intermedia	11.1	20.7
Rural	18.1	12

Note. Excepto para el tamaño de la muestra (*N*) y la edad, los datos se muestran en porcentajes.

La fiabilidad test-retest se estimó para dos franjas de edad (2:6-3:11 y 4:0-7:7) empleando la correlación de Pearson. La media y la desviación típica de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas se presentan en la tabla 5.5. Los coeficientes de estabilidad test-retest promedio para todas las edades se calcularon utilizando la transformación *z* de Fischer. La tabla también recoge las diferencias típicas (tamaño del efecto) entre la primera y la segunda aplicación y los coeficientes de correlación corregidos de acuerdo con la variabilidad de la muestra de tipificación española (Allen y Yen, 2002; Magnusson, 1967). La diferencia típica se calculó usando la diferencia de la puntuación media obtenida en las dos aplicaciones dividida por la desviación típica común (Cohen, 1988).

Table 5.5. Coeficientes de estabilidad de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas (datos españoles)

Todas las edades	Primera aplicación		Segunda aplicación					
Prueba/ punt. de procesamiento/ punt. compuesta	Media	*Dt*	Media	*Dt*	*n*	r_{12}[a]	Corregida *r*[b]	Diferencia típica[c]
Información	10.3	2.2	11.2	2.1	72	.69	.83	.42
Semejanzas	10.3	2.5	10.7	2.7	44	.56	.69	.15
Vocabulario	10.5	2.5	10.8	2.5	45	.76	.83	.12
Comprensión	10.8	2.9	10.4	2.7	45	.70	.72	-.14
Dibujos	11.1	2.6	11.8	3.2	69	.63	.72	.24
Nombres	10.2	2.0	10.9	2.5	72	.63	.84	.31
Cubos	9.8	2.8	10.7	2.8	68	.74	.78	.32
Rompecabezas	10.5	2.4	11.6	2.6	70	.67	.79	.44
Matrices	10.2	2.8	11.7	2.7	43	.70	.74	.55
Conceptos	10.3	2.3	12.1	2.7	44	.52	.72	.72
Reconocimiento	10.9	2.6	11.7	2.7	69	.69	.77	.30
Localización	10.8	2.4	11.0	3.0	70	.58	.73	.07
Búsqueda de animales	10.7	2.2	11.2	2.5	45	.72	.85	.21
Cancelación	10.5	2.5	11.2	2.3	42	.63	.74	.29
Clave de figuras	10.6	2.6	11.0	2.3	44	.71	.78	.16
Cancelación aleat.	10.4	2.5	11.2	2.6	42	.61	.73	.31
Cancelación estruct.	10.2	2.3	11.0	2.2	44	.48	.70	.36
ICV	101.9	11.2	107.0	12.0	72	.67	.82	.44
IVE	100.7	12.3	106.4	13.2	66	.72	.81	.45
IRF	101.2	11.6	111.9	13.4	43	.75	.85	.85
IMT	105.2	11.5	108.3	13.3	67	.66	.80	.25
IVP	103.6	11.3	107.5	11.9	42	.77	.87	.34
CIT	101.7	11.3	109.2	11.7	61	.78	.87	.65
IAV	103.6	10.7	108.2	12.8	69	.66	.83	.39
INV	102.1	11.0	109.6	12.6	60	.74	.86	.63
ICG	100.4	11.3	107.9	12.0	64	.77	.87	.64
ICC	105.5	11.3	109.4	11.6	39	.77	.87	.34

[a] Los coeficientes de estabilidad promedio entre todas las edades se calcularon con la transformación *z* de Fisher.

[b] Se corrigieron las correlaciones de acuerdo con la variabilidad de la muestra de tipificación española (Allen y Yen, 2002; Magnusson, 1967).

[c] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Table 5.5. Coeficientes de estabilidad de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas (datos españoles) *(continuación)*

Edad 2:6–3:11	Primera aplicación		Segunda aplicación					
Prueba/ punt. de procesamiento/ punt. compuesta	Media	*Dt*	Media	*Dt*	*n*	r_{12}	Corregida *r*[a]	Diferencia típica[b]
Información	10.3	2.0	11.4	2.0	27	.76	.89	.55
Dibujos	10.7	2.6	12.0	3.3	27	.77	.83	.44
Nombres	9.8	1.8	11.0	2.3	27	.34	.76	.58
Cubos	9.2	2.7	10.1	2.3	25	.65	.72	.36
Rompecabezas	9.7	2.3	11.1	2.6	26	.68	.81	.57
Reconocimiento	10.9	2.3	12.1	2.7	27	.67	.81	.48
Localización	11.1	2.1	10.6	2.8	25	.32	.66	-.20
ICV	102.7	10.5	109.9	10.6	27	.76	.88	.68
IVE	95.9	10.3	103.4	10.3	24	.57	.80	.73
IMT	106.0	10.6	109.0	12.2	25	.50	.75	.26
CIT	100.3	9.6	109.9	9.4	24	.70	.88	1.01
IAV	101.3	10.1	108.5	13.1	27	.67	.85	.62
INV	101.2	9.3	107.4	10.2	22	.43	.78	.64
ICG	99.3	9.3	108.6	9.7	24	.68	.88	.98

[a] Se corrigieron las correlaciones de acuerdo con la variabilidad de la muestra de tipificación española (Allen y Yen, 2002; Magnusson, 1967).

[b] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Table 5.5. Coeficientes de estabilidad de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas (datos españoles) *(continuación)*

Edad 4:0-7:7	Primera aplicación		Segunda aplicación					
Prueba/ punt. de procesamiento/ punt. compuesta	Media	*Dt*	Media	*Dt*	*n*	r_{12}	Corregida *r* [a]	Diferencia típica [b]
Información	10.2	2.4	11.0	2.2	45	.66	.78	.35
Semejanzas	10.3	2.5	10.7	2.7	44	.56	.69	.15
Vocabulario	10.5	2.5	10.8	2.5	45	.76	.83	.12
Comprensión	10.8	2.9	10.4	2.7	45	.70	.72	-.14
Dibujos	11.3	2.6	11.7	3.1	42	.55	.66	.14
Nombres	10.4	2.2	10.9	2.6	45	.78	.88	.21
Cubos	10.2	2.8	11.1	3.0	43	.77	.80	.31
Rompecabezas	10.9	2.3	11.9	2.6	44	.65	.79	.41
Matrices	10.2	2.8	11.7	2.7	43	.70	.74	.55
Conceptos	10.3	2.3	12.1	2.7	44	.52	.72	.72
Reconocimiento	10.9	2.7	11.4	2.8	42	.71	.76	.18
Localización	10.6	2.6	11.3	3.1	45	.71	.78	.24
Búsqueda de animales	10.7	2.2	11.2	2.5	45	.72	.85	.21
Cancelación	10.5	2.5	11.2	2.3	42	.63	.74	.29
Clave de figuras	10.6	2.6	11.0	2.3	44	.71	.78	.16
Cancelación aleat.	10.4	2.5	11.2	2.6	42	.61	.73	.31
Cancelación estruct.	10.2	2.3	11.0	2.2	44	.48	.70	.36
ICV	101.4	11.7	105.3	12.5	45	.64	.78	.32
IVE	104.0	12.4	108.2	14.4	45	.70	.79	.31
IRF	100.4	12.0	111.9	13.4	45	.60	.74	.90
IMT	103.9	12.5	107.8	14.0	45	.66	.76	.29
IVP	103.0	11.2	107.5	11.9	45	.67	.82	.39
CIT	102.6	11.6	108.7	13.0	45	.76	.86	.50
IAV	104.6	10.7	108.1	12.7	45	.62	.81	.30
INV	102.8	11.0	110.9	13.8	45	.78	.88	.65
ICG	101.2	12.1	107.6	13.2	45	.73	.83	.51
ICC	104.5	11.0	109.4	11.6	45	.69	.83	.43

[a] Se corrigieron las correlaciones de acuerdo con la variabilidad de la muestra de tipificación española (Allen y Yen, 2002; Magnusson, 1967).

[b] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Como se desprende de los datos expuestos en la tabla 5.5, las puntuaciones de la WPPSI-IV muestran una adecuada consistencia temporal para las dos franjas de edad. Los coeficientes de estabilidad promedios corregidos para *Información*, *Vocabulario*, *Nombres* y *Búsqueda de animales* son buenos (alrededor de .80); y para el resto de las pruebas son aceptables (alrededor de .70). El coeficiente de estabilidad promedio corregido para la puntuación de procesamiento Cancelación aleatoria es .73 y para la puntuación de procesamiento Cancelación estructurada es .70. Los coeficientes de los índices primarios son buenos o excelentes, entre .80 (IMT) y .87 (IVP). El coeficiente del CI total es excelente (.87). Los coeficientes de los índices secundarios son buenos o excelentes, entre .83 (IAV) y .87 (ICC).

Tal y como indican los datos de la tabla 5.5, en conjunto, las puntuaciones medias de todas las pruebas fueron superiores en la segunda aplicación (retest) que las obtenidas en la primera aplicación (test); con dos excepciones en distintas franjas de edad: la puntuación media de *Localización* correspondiente a la franja de edad de 2:6 a 3:11 años es ligeramente inferior en la segunda aplicación, y lo mismo sucede con la puntuación media de *Comprensión* correspondiente a la franja de edad de 4:0 a 7:7 años. En general, este incremento de puntuación en la segunda aplicación es menor en las pruebas de Comprensión verbal y de Memoria de trabajo que en las

de Razonamiento perceptivo y Velocidad de procesamiento. Estos resultados son consistentes con los obtenidos en otras escalas de inteligencia Wechsler (Wechsler, 2002, 2003, 2008).

Acuerdo entre correctores

Durante la fase de tipificación española, una muestra de 150 protocolos de la WPPSI-IV se sometieron a una doble corrección por parte de dos correctores independientes. Posteriormente se comprobó el grado de acuerdo entre ellos. Dado que los criterios de corrección de la mayoría de las pruebas eran sencillos y objetivos, el acuerdo entre correctores fue alto, de .95 a .99.

En la tabla 5.4. se recogen las características demográficas de los niños de la muestra a partir de la cual se valoró el acuerdo entre correctores, 48 niños con edades de 2:6 a 3:11 años y 102 niños con edades de 4:0 a 7:7 años. En ella se indican la edad media y los porcentajes de representación por sexo, nivel educativo de los padres, zona geográfica y tipo de población.

Los 150 protocolos fueron corregidos por dos correctores con un adecuado nivel de cualificación profesional. Ninguno de los correctores tenía experiencia previa con el sistema de corrección de la WPPSI-IV. Los coeficientes de fiabilidad entre correctores se calcularon de acuerdo con los métodos de correlación intraclase (McGraw y Wong, 1996; Shrout y Fleiss, 1979). En este análisis se utilizaron las puntuaciones totales basadas en ítems corregidos subjetivamente. Los coeficientes de fiabilidad entre correctores fueron los siguientes: .99 para *Información* (n = 150), .95 para *Semejanzas* (n = 102), .95 para *Vocabulario* (n = 102), .95 para *Comprensión* (n = 102) y .99 para *Nombres* (n = 150). Estos resultados muestran que, aunque la corrección de estas pruebas requiera una valoración más subjetiva, pueden ser puntuadas de modo fiable, incluso por correctores sin experiencia previa con los criterios de corrección de la WPPSI-IV.

Diferencias entre puntuaciones

La diferencia entre dos puntuaciones obtenidas en una única aplicación es una comparación intraindividual, y esta diferencia debe tenerse en cuenta en la interpretación. Dado que la comparación se basa en el propio rendimiento del niño, al evaluar la diferencia entre puntuaciones, debe considerarse el nivel de aptitud general del niño. Por ejemplo, para un niño con una aptitud general alta, una puntuación alta como valor absoluto (muy por encima de 100) en un índice primario puede representar un punto débil relativo respecto a su rendimiento general o a su rendimiento en otro dominio cognitivo. En cuanto a las pruebas, también debe considerarse el nivel de aptitud general del niño. Por ejemplo, para un niño con una aptitud general alta, un resultado alto (muy por encima de 10) en una puntuación escalar de una prueba puede representar un punto débil relativo respecto a su rendimiento general o a su rendimiento en otra prueba. Por el contrario, para un niño con una aptitud general limitada, un resultado bastante bajo en una puntuación de un índice primario o de una prueba puede indicar un punto fuerte relativo respecto a su rendimiento general o su rendimiento en otros dominios cognitivos.

A la hora de interpretar los resultados obtenidos por un niño en la WPPSI-IV es importante analizar las diferencias que se observan entre las distintas puntuaciones. En este sentido hay dos aspectos relevantes: la significación estadística de la diferencia, determinada por su comparación con los valores críticos; y la frecuencia (tasa base) en que esa diferencia se ha observado en la población. Estos aspectos han de responder dos preguntas, respectivamente: esa diferencia *¿es real o se debe al error de medida? y ¿es clínicamente significativa?*

Significación estadística de las diferencias entre puntuaciones

Una diferenciación estadísticamente significativa entre dos puntuaciones, basada en su comparación con los valores críticos, indica la probabilidad de que esta diferencia se obtenga por azar (p. ej., $p < .05$ o .10), si la diferencia verdadera entre puntuaciones es 0 (Matarazzo y Herman, 1985). El nivel de significación refleja el grado de confianza que el profesional puede tener respecto a que la diferencia observada entre puntuaciones es una diferencia verdadera; y determina qué diferencia debe haber entre dos puntuaciones (valor crítico) para alcanzar un nivel de

confianza específico que indique que la diferencia es una diferencia verdadera y no debida al azar. La elección del nivel de significación del valor crítico se basa en diversos factores, como el objetivo de la evaluación (p. ej., describir las aptitudes cognitivas ante un diagnóstico diferencial de un determinado trastorno del aprendizaje) o el número de comparaciones de puntuaciones que pueden llevarse a cabo al mismo tiempo. La WPPSI-IV ofrece valores críticos para evaluar la diferencia entre puntuaciones con los niveles de significación .01, .05., .10 y .15, por grupo y franja de edad. El hecho de disponer de más niveles de significación permite que el profesional adecue la selección de los valores críticos a la situación clínica concreta y, con ello, pueda lograr un equilibrio entre la potencia estadística y el posible error de tipo I (falsos positivos).

La elección de un nivel de significación menos estricto (p. ej., .15) puede ser adecuada cuando se van a comparar pocas puntuaciones o en situaciones relativamente de bajo riesgo, como cuando el motivo de consulta es obtener el rendimiento medio de un niño a fin de conocer mejor los puntos fuertes y débiles del niño. Por el contrario, cuando se van a comparar muchas puntuaciones al mismo tiempo (p. ej., comparaciones entre todas las puntuaciones de los índices primarios) o en situaciones de alto riesgo, debe elegirse un nivel de significación más estricto (p. ej., .01 o .05). Dado que los niveles de significación más estrictos dan lugar a menos diferencias significativas, es importante considerar tanto el motivo de evaluación como el número de comparaciones simultáneas de puntuaciones que se van a realizar, antes de elegir el nivel de significación de los valores críticos. Si el error de tipo I es un aspecto crítico o si se prevé realizar muchas comparaciones de puntuaciones, el profesional puede optar por dar más relevancia a la evaluación de la tasa base de la diferencia entre puntuaciones de la muestra de tipificación que a la significación de la diferencia entre puntuaciones.

Valores críticos corregidos y no corregidos

Los valores críticos de la WPPSI-IV se calcularon tomando en cuenta la varianza compartida entre las puntuaciones comparadas. Se utilizan los valores críticos corregidos cuando las puntuaciones que se comparan comparten contenido del test, como cuando se compara una puntuación de una prueba de un índice primario con la media de las puntuaciones de los índices primarios (MIP) o cuando se compara la puntuación de una prueba principal para obtener el CI total con la media de las puntuaciones escalares de las pruebas que conforman el CI total. En cambio, se utilizan los valores críticos no corregidos cuando se comparan puntuaciones que no comparten contenido del test, como las comparaciones entre índices primarios o puntuaciones de pruebas.

Puntos fuertes y débiles

La comparación de un índice primario o una puntuación de una prueba con una estimación de la aptitud general, como el MIP o la media de las puntuaciones de las pruebas que conforman el CI total, ofrece información relevante sobre los puntos fuertes y débiles intraindividuales en relación con una estimación global del rendimiento en varios dominios de aptitud cognitivos (Grégoire et al., 2011; Longman, 2004; Naglieri y Paolitto, 2005; Wechsler, 2002, 2003, 2008). Una puntuación de un índice primario que sea significativamente superior a la MIP obtenida por el mismo sujeto puede reflejar un punto fuerte relativo, mientras que si es significativamente inferior a la MIP puede indicar un punto débil relativo. Del mismo modo, una puntuación de una prueba que se desvía significativamente de la media de las puntuaciones de las pruebas que conforman el CI total puede indicar un punto fuerte o débil en relación con el rendimiento general del niño.

En la tabla B.1 del *Manual de aplicación y corrección* figuran los valores críticos corregidos para comparar cada uno de los índices primarios con la MIP o el CI total. Los valores críticos corregidos, que se utilizan para comparar cada uno de los índices primarios y la MIP, se han calculado con la fórmula de Davis (1959):

$$z\sqrt{\left(\sum_{j}^{i} ETM_{a}^{2}\right)/K^{2} + \left[\frac{K-2}{K}\right]\left(ETM_{i}^{2}\right)}$$

donde z es el valor de la curva normal asociado con el nivel de significación bilateral requerido, obtenido con la corrección propuesta por Dunn-Šidák para las comparaciones múltiples; K es el número de índices que se comparan; $\sum_{j}^{i} ETM_{a}^{2}$ es la suma del cuadrado de los ETM de todos los índices que se comparan; y ETM_{i}^{2} es el cuadrado del ETM del índice que se analiza.

Dado que algunas de las pruebas que conforman los índices primarios también contribuyen al CI total, los valores críticos para comparar cada uno de los índices primarios y el CI total se obtienen *mediante el error típico de la diferencia* (ET_{dif}), *que se calcula con la fórmula siguiente:*

$$ET_{dif} = \sqrt{ETM_{xi}^{2} + ETM_{xj}^{2} - 2\sqrt{\frac{\sum ETM_{i}^{2}}{\sum ETM_{j}^{2}}}\left(ETM_{xi}\right)\left(ETM_{xj}\right)}$$

donde ETM_{xi} y ETM_{xj} son los errores típicos de medida del índice y del CI total, $\sum ETM_{i}^{2}$ es la suma del cuadrado de los errores típicos de medida de todas las pruebas compartidas por el índice y el CI total, y $\sum ETM_{j}^{2}$ es la suma del cuadrado de los errores típicos de medida de todas las pruebas que conforman el CI total.

Al multiplicar el ET_{dif} por un valor z apropiado se obtiene el valor crítico, es decir, la diferencia requerida para la significación estadística en cualquier nivel de confianza. Se utilizó la misma metodología para calcular los valores críticos para las diferencias entre el ICG y el CI total. Estos valores críticos figuran en la tabla C.4 de este manual.

La tabla B.3 del *Manual de aplicación y corrección* presenta los valores críticos para comparar cada una de las pruebas que conforman los índices primarios y la media de las puntuaciones escalares de las pruebas que conforman los índices primarios (MPE-IP) o la media de las puntuaciones escalares de las pruebas que conforman el CI total (MPE-CI). Se ofrecen los valores críticos corregidos para las comparaciones entre las pruebas de los índices primarios y la MPE-IP, así como entre las pruebas del CI total y la MPE-CI, a fin de tener en cuenta el contenido compartido por las pruebas. Estos valores críticos se calcularon mediante la misma fórmula usada para las comparaciones entre cada uno de los índices primarios y la MIP, utilizando los valores del ETM de la prueba en vez de los valores del índice cuando correspondía. Los valores críticos para comparar las pruebas de los índices primarios que no contribuyen al CI total (p. ej., *Rompecabezas* y *Localización*) con la MPE-CI, se calcularon mediante la fórmula siguiente sin ajuste de la varianza compartida:

$$z\sqrt{\left(\sum_{j}^{i} ETM_{a}^{2}\right)/K^{2} + ETM_{i}^{2}}$$

donde z es el valor de la curva normal asociado con el nivel de significación bilateral requerido, calculado con la corrección propuesta por Dunn-Šidák para las comparaciones múltiples; $\sum_{j}^{i} ETM_{a}^{2}$ es la suma del cuadrado de los ETM de las pruebas que conforman el CI total; ETM_{i}^{2} es el cuadrado de los ETM de la prueba que se analiza; y K es el número de pruebas que contribuyen al CI total.

Dado que los valores críticos para realizar comparaciones con la MPE-CI que figuran en la tabla B.3 se basan en las pruebas principales para obtener el CI total, debe comprobarse si se ha realizado alguna sustitución, de las permitidas, para obtener el CI total, y si se utiliza la MPE-CI como puntuación de comparación para las pruebas. En tal caso, debe considerarse el valor crítico y la tasa base de la prueba principal en vez del valor crítico y la tasa base de la prueba sustituida. Por ejemplo, si *Rompecabezas* sustituyó a *Cubos* para obtener el CI total (y, por tanto, *Rompecabezas* contribuyó a la MPE-CI), debe considerarse el valor crítico y la tasa base de *Cubos* (la prueba principal que se sustituyó) en vez del valor crítico y la tasa base de *Rompecabezas*. Se sigue este proceso aparentemente contradictorio porque los valores críticos y las tasas bases de la prueba principal (p. ej., *Cubos*) tienen en cuenta el contenido compartido entre las puntuaciones, mientras que los valores de la prueba complementaria (p. ej., *Rompecabezas*) no los tienen en cuenta. Esta observación también figura en la sección «Cumplimentar el cuadernillo de anotación» del capítulo 2 del *Manual de aplicación y corrección*.

Comparación entre índices o pruebas

Comparar las diferencias entre índices o pruebas permite que el profesional pueda evaluar la variabilidad del rendimiento intraindividual en términos de la diferencia entre las puntuaciones de dos índices o de dos pruebas. Los valores críticos para comparar las puntuaciones de dos índices o de dos pruebas figuran en las tablas B.5 y B.7 del *Manual de aplicación y corrección*, y se calcularon mediante la fórmula siguiente:

$$z\sqrt{ETM_a^2 + ETM_b^2}$$

donde z es el valor de la curva normal asociado con el nivel de significación bilateral requerido, y ETM_a y ETM_b son los errores típicos de medida de las dos puntuaciones. Se utilizó la misma metodología para calcular los valores críticos para las diferencias entre el ICG y el ICC. Estos valores críticos figuran en la tabla C.4 de este manual.

Tasas base de las diferencias entre puntuaciones

La prevalencia o frecuencia de una diferencia observada en la población general se denomina *tasa base*. A menudo, la diferencia entre dos puntuaciones es estadísticamente significativa aun siendo frecuente. Por consiguiente, la significación estadística de las diferencias entre puntuaciones y la frecuencia de las diferencias tienen implicaciones distintas para la interpretación de los resultados de un test. Para más información sobre la significación, estadística y clínica, de las diferencias entre puntuaciones, véanse los trabajos de Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Matarazzo y Herman, 1985; Sattler, 2008a; y Silverstein, 1981.

Dado que la frecuencia de las diferencias entre puntuaciones varía según el nivel de aptitud y no de la edad (Grégoire et al., 2011), en las tablas B.2 y B.6 del *Manual de aplicación y corrección* y en la tabla C.5 de este manual se presentan las tasas base para las comparaciones entre puntuaciones de los índices para la muestra total y por nivel de aptitud (según rangos de puntuación determinados del CI total o del ICG). Por ejemplo, de los niños con un CI total igual o inferior a 79, un 10% obtuvieron un ICV al menos 12.6 puntos inferior a su MIP, mientras que el 10% obtuvieron un ICV al menos 7.8 puntos superior a su MIP. Puesto que la frecuencia de las diferencias también puede variar según la dirección de la diferencia (Sattler, 2008a), las tasas bases para las comparaciones entre puntuaciones de índices y pruebas (tablas B.2, B.4, B.6 y B.8 del *Manual de aplicación y corrección* y la tabla C.5 de este manual) se ofrecen según la dirección de la diferencia (positiva o negativa).

Análisis de dispersión entre pruebas

La dispersión de las puntuaciones escalares refleja la variabilidad del rendimiento del niño en las diferentes pruebas (Matarazzo, Daniel, Prifitera y Herman, 1988; Sattler, 2008a). Esta variabilidad a menudo se ha considerado importante para realizar un diagnóstico. Aunque se han propuesto varias medidas de la dispersión (véase Schinka, Vanderploeg y Curtiss, 1997), el índice de dispersión utilizado en el *Manual de aplicación y corrección* se calcula a partir de la diferencia entre la puntuación escalar más alta y la puntuación más baja obtenida por el sujeto en las pruebas. Antes de analizar la dispersión, debería considerarse la frecuencia de dicha dispersión en la muestra de tipificación. Los porcentajes acumulados de la dispersión entre las pruebas que conforman los índices primarios y entre las pruebas que permiten obtener el CI total figuran en la tabla B.9 del *Manual de aplicación y corrección* para cada franja de edad. Por ejemplo, entre las diez pruebas que conforman los índices primarios, el 8.7 % de los niños de 4:00 a 7:00 años de la muestra de tipificación española obtuvieron una dispersión de 11 puntos o más; y, entre las seis pruebas que permiten obtener el CI total, el 2.4 % de los mismos niños obtuvieron una dispersión de 11 puntos o más. Para aquellos casos poco habituales en que el niño obtiene la puntuación más alta y la más baja en pruebas que contribuyen al mismo índice (p. ej., puntuación más alta en *Información* y más baja en *Semejanzas*), puede ser preferible considerar la tasa base de la diferencia entre las pruebas correspondientes, puesto que dichas tasas bases se ofrecen según la dirección de la diferencia y proporcionan datos más descriptivos a efectos de interpretación.

Resumen

En este capítulo se han presentado los estudios de fiabilidad realizados sobre la WPPSI-IV. Las tablas citadas ofrecen los datos necesarios para interpretar de forma fiable las diferencias entre las puntuaciones obtenidas por un niño en la WPPSI-IV. Puede suceder, como ya se ha visto, que diferencias aparentemente grandes entre dos puntuaciones, por ejemplo, no resulten estadísticamente significativas y, por el contrario, que diferencias estadísticamente significativas sean frecuentes en la población general. El profesional debería tener presente que el error de medida contribuye a las diferencias observadas y que la presencia de ese error será, generalmente, menor en las diferencias entre puntuaciones compuestas (p. ej., entre el ICV y el IRF) que en las diferencias entre puntuaciones de las pruebas. Esto sucede porque, por lo general, las puntuaciones compuestas son más fiables que las de las pruebas. Como siempre, cuando se interpreta una puntuación, debe combinarse la información relevante procedente de diversas fuentes, incluidos el motivo de evaluación; el historial médico, académico o psicosocial del niño; las puntuaciones obtenidas en otros tests (p. ej., medidas del estado mental, de la conducta, del rendimiento académico y del funcionamiento ejecutivo); y las propiedades estadísticas de los tests utilizados durante la evaluación.



Validez

La validez es el aspecto más importante desde el punto de vista de la construcción de un test (AERA, 1999; Angoff, 1988; Sattler, 2008a). Tradicionalmente, los investigadores y constructores de tests se han referido a tres principales tipos de validez: validez de contenido, validez en relación con un criterio y validez de constructo. Así, se consideraba que un test tenía validez de contenido si representaba adecuadamente los aspectos relevantes del constructo que se estaba midiendo. Un test poseía validez en relación con un criterio si las puntuaciones guardaban relación con un criterio específico externo, como el rendimiento en alguna otra medida o la pertenencia a un grupo. Por último, la validez de constructo de un test se conseguía si realmente se medía el significado del constructo que se pretendía medir con el test. La evidencia de validez del constructo puede proceder de diferentes fuentes, como el análisis factorial, la revisión por parte de expertos, los estudios multirrasgo-multimétodo y las investigaciones clínicas.

Las definiciones actuales de validez son de carácter más unitario. Como se afirma en *Standards for educational and psychological testing*, «en el desarrollo del concepto de [...] validez no se habla ya de diferentes tipos de validez, sino de diferentes fuentes de evidencia de validez, todas ellas encaminadas a proporcionar una información relevante para lograr una interpretación específica de las puntuaciones del test» (AERA, 1999, p. 5). La validez, por tanto, se refiere al grado en el que existe evidencia que apoye la interpretación de las puntuaciones del test con un fin determinado. Por consiguiente, en el examen de la validez de un test debe intervenir la valoración del propio usuario del test.

Aunque los constructores de los tests son responsables de proporcionar evidencias iniciales y empíricas de validez, el profesional es quien debe evaluar si estas evidencias justifican el uso que el profesional desea realizar del test con un fin determinado. Esto no significa que el test sea válido únicamente en los contextos y para los propósitos establecidos por los constructores del test. El examen exhaustivo de los datos sobre la validez de una escala también ha de incluir el de los trabajos de investigación más importantes acerca de las versiones anteriores de la escala, así como el análisis de los estudios que aplican la última versión de la escala en diferentes contextos, en poblaciones distintas y con propósitos diversos. De esta forma, se espera que el uso de la WPPSI-IV proporcione en el futuro más información sobre la validez de la escala. Si se desea más información sobre los datos que apoyan la validez de la predecesora de la WPPSI-IV, la WPPSI-III, véanse Flanagan y Harrison (2012), Lichtenberger y Kaufman (2003) y Wechsler (2002).

Este capítulo presenta diferentes fuentes de evidencias de la validez de la WPPSI-IV que apoyan el uso de la escala como medida de la aptitud intelectual, incluidos los datos sobre el contenido, los procesos de respuesta, la estructura interna de la escala y su relación con otras variables.

Evidencia basada en el contenido

El estudio de la relación entre el contenido de un test y el constructo que se pretende medir constituye una importante fuente de información acerca de la validez del test. Los datos que prueban la validez de contenido de un test no se basan en análisis estadísticos o en pruebas empíricas, sino que se fundamentan en el grado en que los ítems del test se relacionan y representan adecuadamente el rasgo o función que se pretende medir. El contenido de un test también implica la redacción y el formato de los ítems, así como los procedimientos para la aplicación y corrección del test.

El objetivo de la revisión del contenido de la WPPSI-IV fue asegurar que todos los ítems y pruebas abordaran adecuadamente los distintos dominios del funcionamiento intelectual que la escala pretende evaluar. Los ítems y las pruebas cubren un amplio rango de procesos cognitivos, incluidos el razonamiento verbal y analógico, la formación de conceptos, el procesamiento secuencial y simultáneo, el procesamiento visoespacial, la comprensión auditiva, la flexibilidad cognitiva, la memoria de trabajo y la velocidad de procesamiento psicomotora (Lichtenberger y Kaufman, 2003; Sattler, 2008a). Para mejorar la amplitud y la relevancia del contenido, se llevaron a cabo revisiones de la literatura exhaustivas y se consultó con expertos para examinar el contenido de la WPPSI-IV y evaluar así la pertinencia de incorporar nuevos ítems y pruebas. El capítulo 2 de este manual describe con detalle el contenido y la estructura de la WPPSI-IV.

Evidencia basada en los procesos de respuesta

En la WPPSI-IV, la evidencia de validez basada en los procesos de respuesta debería confirmar que el niño usa el proceso cognitivo esperado al realizar las tareas de una prueba determinada. Este tipo de evidencia puede proceder de fuentes teóricas o de análisis psicométricos.

La mayoría de las pruebas de la WPPSI-IV cuentan con una fuerte base teórica y empírica de validez, basada en los procesos de respuesta. Así, las pruebas que se han conservado de versiones anteriores cuentan con el apoyo de numerosas investigaciones sobre los procesos de respuesta que intervienen en la ejecución de una tarea (Carroll, 1993; Lichtenberger y Kaufman, 2003; Sattler, 2008a). A fin de aportar más datos sobre la validez de los procesos de respuesta que participan en las pruebas nuevas, se revisó la literatura científica, se consultó con expertos y se realizaron exámenes psicométricos. Los detalles sobre las bases teóricas de todas las pruebas, así como la información acerca de los procesos de respuesta relacionados, se ofrecen en el capítulo 2 de este manual.

Los análisis empíricos y cualitativos que se realizaron durante todo el proceso de desarrollo de la escala permitieron obtener nuevas evidencias sobre la validez de los procesos de respuesta. Por ejemplo, se evaluaron las frecuencias de respuesta a los ítems con respuesta múltiple, para identificar qué respuestas incorrectas se daban con más frecuencia. Las respuestas incorrectas más frecuentes se examinaron para determinar la posibilidad de considerarlas como respuestas aceptables o posibles, y se pidió a los niños que explicaran los criterios que los habían llevado a dar tales respuestas. En todas las fases del proceso de desarrollo de la WPPSI-IV, se preguntó a los niños qué criterios habían seguido para agrupar las imágenes de la prueba *Conceptos*. Al combinar las respuestas de los niños con las frecuencias de dichas respuestas, las propias respuestas literales aportaron los criterios para indicar las respuestas incorrectas y correctas de los niños. Si de este análisis se concluía que existía la posibilidad de que el niño diera una respuesta inesperada, pero aceptable, se modificaba el contenido de los ítems, y estos se valoraban de nuevo en otra fase del proceso de desarrollo o se eliminaban. Asimismo se pidió a los niños que explicaran las estrategias que habían seguido para resolver los ítems de *Matrices*, y, en función de sus respuestas, se modificaron las consignas y los distractores de la prueba o se eliminó el ítem.

Las preguntas directas al niño también son útiles para poner de manifiesto diferencias en los procesos de respuesta en las pruebas de un mismo dominio e incluso en las pruebas más tradicionales que cuentan con un largo historial de uso. Así, para la prueba *Cubos*, se pidió a los niños que explicaran cómo habían utilizado los cubos de dos colores en la parte B. En atención a las respuestas se añadió un nuevo ítem de ejemplo para garantizar que los niños eran conscientes de que los cubos de la parte B tenían caras de uno y dos colores y que los dos colores estaban separados por una línea en diagonal. Las aplicaciones y las entrevistas llevadas a cabo en los estudios piloto con niños pequeños que muestran una cierta desconfianza o que presentan retraso en el desarrollo del lenguaje indicaron que las aptitudes medidas por *Semejanzas* y *Comprensión* están presentes en dichos niños aunque no siempre pueden expresar verbalmente sus respuestas a los primeros ítems. Por ello se consideró conveniente que los primeros ítems de estas pruebas fueran gráficos. Los ítems gráficos de *Semejanzas* evolucionaron a partir de los intentos de mejorar el suelo de los ítems existentes y en función de las respuestas de los niños sobre la simplificación de las consignas. Se evidenció que a los niños les era de ayuda que se les demostrara cómo extraer la

categoría superior a la que pertenecían dos elementos, puesto que de este modo podían responder correctamente, mientras que sin tal demostración no lo hacían. Por consiguiente se crearon los ítems gráficos de *Semejanzas*.

A lo largo del desarrollo de las pruebas nuevas, para obtener más información sobre los procesos de respuesta en los niños pequeños, se siguió un procedimiento iterativo basado en la observación, la pregunta directa y la modificación de los requisitos de la tarea, el modo de estímulo y el modo de respuesta. Como resultado de este procedimiento, la prueba *Reconocimiento* pasó de ser una tarea de secuenciación a una tarea en la que interviene la interferencia proactiva. Inicialmente, esta prueba consistía en que el niño debía señalar las imágenes en el mismo orden en que se le habían mostrado. No obstante, se evidenció que a veces incluso los niños muy brillantes no realizaban la secuenciación pese a haber completado varios ítems de demostración y de ejemplo. Tras consultar con especialistas en memoria de trabajo en la infancia y revisar la literatura pertinente, se concluyó que los niños pequeños frecuentemente no muestran estrategias de repetición y que, por lo tanto, las tareas de secuenciación no eran adecuadas a su desarrollo. Por ello, se modificó la tarea para que se basara en la interferencia proactiva e incluyese las demandas cognitivas concurrentes que requieren la intervención de la memoria de trabajo.

En las primeras fases de la investigación, cuando los estímulos de *Localización* se mostraban mediante un cuaderno de estímulos en vez de mediante tarjetas y planos, tres de cada cuatro niños de 3 años a los que se les aplicó la prueba no respondieron correctamente a ningún ítem. Cuando se les preguntó a los niños cómo habían resuelto los ítems, se constató que los niños no habían relacionado los dibujos estáticos mostrados con el hecho de dar una respuesta utilizando las tarjetas; por ello, respondían colocando las tarjetas donde querían. Consecuentemente, *Localización* se modificó a fin de que se usara el mismo método para presentar los estímulos de los animales que para responder; es decir, las tarjetas.

Para la prueba *Búsqueda de animales*, tras ser sometida a un análisis similar del proceso de respuesta, descrito en el capítulo 2 de este manual, se determinó que cada fila de grupos de búsqueda fuera de un color diferente y que se eliminara la situación de no coincidencia. El análisis del proceso de respuesta también determinó la formulación de las consignas de *Cancelación* («cosas que la gente se pone» en vez de «prendas de vestir») a fin de garantizar que se midiera la velocidad de procesamiento más que el vocabulario receptivo. Al igual que sucede con otras evidencias de validez, se espera que las futuras investigaciones ofrezcan más datos sobre los procesos de respuesta que implican los ítems y las pruebas de la WPPSI-IV.

Evidencia basada en la estructura interna

El estudio de la estructura interna de un test «puede indicar el grado en el que las relaciones entre los ítems y componentes del test se ajustan al constructo en el que se basa la interpretación de la puntuación del test» (AERA, 1999). Los estudios de investigación que han utilizado las escalas de inteligencia de Wechsler han aportado fuertes evidencias de validez basadas en la estructura interna de la escala (Wechsler, 2002, 2003, 2008).

El tipo de evidencias aportadas ha evolucionado paralelamente a los avances experimentados en la teoría y en la medición de la inteligencia, así como en los campos de la neuropsicología y el desarrollo cognitivo. En una serie de artículos (1957a, 1957b, 1959), Cohen presentó los resultados del análisis factorial de la escala Wechsler-Bellevue, la WAIS y la WISC, y sugirió la presencia de un tercer factor débil dentro de la escala manipulativa. Así, para confirmar la existencia de este tercer factor, se incluyó la prueba *Búsqueda de símbolos* en la *Wechsler Intelligence Scale for Children-Third Edition* (WISC-III; Wechsler, 1991) pero se puso de manifiesto la presencia de un cuarto factor: Velocidad de procesamiento. Del mismo modo, la inclusión de la prueba *Letras y números* en la *Escala de inteligencia de Wechsler para adultos* (WAIS-III; Wechsler, 1997, 1999) también condujo a cuatro factores. Consecuentemente, se desarrollaron cuatro índices para la WISC-III y cuatro índices para la WAIS-III con el objetivo de representar aptitudes cognitivas específicas: Comprensión verbal, Organización perceptiva, Independencia de la distracción (o Memoria de trabajo en la WAIS-III) y Velocidad de proceso; manteniendo al mismo tiempo

las puntuaciones tradicionales para el CI verbal, el CI manipulativo y el CI total. Conforme los nuevos índices ganaban aceptación y un mayor apoyo empírico, aumentaron las evidencias de la utilidad clínica de la escala. Con la incorporación de *Claves* y *Búsqueda de símbolos* en la WPPSI-III para los niños de 4:0 a 7:3 años, se puso de manifiesto la presencia de un factor de Velocidad de procesamiento junto con los factores verbal y manipulativo, de lo que resultó un modelo de tres factores para las edades mencionadas.

La WISC-IV y la WAIS-IV añadieron nuevas medidas de razonamiento fluido, memoria de trabajo y velocidad de procesamiento. Para ambas escalas se mantuvo la estructura de cuatro factores, se cambió la denominación *Índice de organización perceptiva* por *Índice de razonamiento perceptivo* para reflejar el mayor énfasis del factor en el razonamiento fluido, y se eliminaron el CIV y el CIM. Estas modificaciones son consistentes con la teoría de la inteligencia contemporánea, los estudios de análisis factorial y las investigaciones clínicas. Además, para la WISC-IV, se cambió la denominación *Índice de independencia de la distracción* por *Índice de memoria de trabajo*. Diversos estudios independientes han confirmado la estructura de cuatro factores de la WISC-IV y de la WAIS-IV (p. ej., Bodin, Pardini, Burns y Stevens, 2009; Ward, Bergman y Hebert, 2011; Watkins, 2010; Watkins, Wilson, Kotz, Carbone y Babula, 2006).

Desde la publicación de la WISC-IV y de la WAIS-IV, los investigadores han identificado una estructura factorial alternativa, común a ambas escalas, que también tiene ventajas. Los análisis factoriales de varios estudios sugieren la presencia de cinco factores en dichas escalas (Benson et al., 2010; Keith et al., 2006; Weiss et al., 2013; Weiss et al., 2013). Los resultados de estos estudios muestran que los factores Comprensión verbal y Velocidad de procesamiento son iguales a los de los tests publicados, y que las pruebas de los otros dos factores (Razonamiento perceptivo y Memoria de trabajo) se distribuyen en tres factores, que representan las aptitudes visoespaciales, de razonamiento fluido y de memoria de trabajo. En la WISC-IV, las pruebas *Cubos* y *Figuras incompletas* contribuyen al factor Visoespacial; *Conceptos, Matrices* y *Aritmética* contribuyen al factor Razonamiento fluido; y *Dígitos* y *Letras y números* contribuyen al factor Memoria de trabajo. En la WAIS-IV, el modelo de cinco factores se pone de manifiesto cuando todas las pruebas están disponibles para las edades de 16 a 69 años, incluidas *Balanzas, Letras y números* y *Cancelación*. Por lo general, las pruebas *Cubos, Puzles visuales* y *Figuras incompletas* contribuyen al factor Visoespacial; *Matrices, Balanzas* y *Aritmética* contribuyen al factor Razonamiento fluido; y *Dígitos* y *Letras y números* contribuyen al factor Memoria de trabajo.

La estructura interna de la WPPSI-IV se analizó teniendo en cuenta los datos aportados por los análisis factoriales; la mayor amplitud del constructo, que incluye la memoria de trabajo (*Reconocimiento* y *Localización*) para las edades de 2:6 a 7:7 años; y la incorporación de pruebas de Velocidad de procesamiento más adecuadas al desarrollo de los niños de 4:0 a 7:7 años (*Búsqueda de animales, Cancelación* y *Clave de figuras*). Se esperaba que los resultados del análisis factorial indicaran la presencia de tres factores para las edades de 2:6 a 3:11 años (Comprensión verbal, Visoespacial y Memoria de trabajo) y de cinco factores para las edades de 4:0 a 7:7 años (Comprensión verbal, Visoespacial, Razonamiento fluido, Memoria de trabajo y Velocidad de procesamiento). El análisis de la estructura interna de la WPPSI-IV conllevó el examen de las relaciones entre los ítems, las pruebas y las puntuaciones compuestas para aportar evidencias de la validez de la escala. En el capítulo 5 de este manual se describe la consistencia interna de cada una de las pruebas.

Estudios de intercorrelación

Campbell y Fiske (1959) presentaron una metodología teórica para interpretar los patrones de correlaciones en una matriz multirrasgo-multimétodo con el objetivo de demostrar evidencias de *validez convergente* y *validez discriminante*. Este estudio, constantemente citado, es una de los más importantes sobre la validez de los tests (Eid y Nussbeck, 2009). Su metodología se basaba en el estudio del patrón de correlaciones de una matriz para la que se esperan correlaciones relativamente altas (validez convergente) para algunas variables y correlaciones relativamente bajas (validez discriminante) para otras variables. Los datos que apoyan las hipótesis formuladas a priori sobre el patrón de relaciones pueden interpretarse como una evidencia de validez de constructo.

Intercorrelaciones de pruebas, puntuaciones de procesamiento y puntuaciones compuestas

Para evaluar la validez de la WPPSI-IV, se formularon varias hipótesis respecto a los estudios de intercorrelación. En primer lugar, se esperaba que todas las puntuaciones mostrasen como *mínimo correlaciones de bajas a moderadas, hipótesis basada en la suposición de que todas las* pruebas miden un factor general de inteligencia (es decir, *g*).

En segundo lugar, se esperaba que las pruebas de un dominio cognitivo específico (p. ej., Comprensión verbal, Visoespacial y Memoria de trabajo) tuviesen correlaciones más altas entre ellas que con las pruebas de otros dominios.

En tercer lugar, los resultados obtenidos en estudios previos indican que algunas pruebas están más relacionadas que otras con el factor *g*. Por ejemplo, *Información, Semejanzas, Vocabulario, Comprensión, Dibujos, Nombres, Cubos, Rompecabezas* y *Matrices* mostraron tener saturaciones altas en *g* (Sattler, 2008a; Wechsler, 1989, 2002). A partir de estos datos se realizaron varias suposiciones. Así, se esperaba que las pruebas con saturaciones altas en *g*, independientemente de la escala de la que forman parte, tuvieran correlaciones relativamente altas entre ellas. Del mismo modo, se esperaba que la correlación entre dos pruebas que presentan saturaciones elevadas en *g* y, además, forman parte del mismo dominio cognitivo (p. ej., Comprensión verbal) fuera más alta que la correlación entre dos pruebas que, aunque muestren saturaciones elevadas en *g*, forman parte de distintos dominios cognitivos (p. ej., Comprensión verbal y Visoespacial).

En cuarto lugar, y a partir de estudios anteriores (Wechsler, 2002), se esperaba que *Conceptos* mostrara correlaciones con las pruebas de Comprensión verbal y de Razonamiento fluido y que *Dibujos* mostrara correlaciones con las pruebas de Visoespacial y de Razonamiento fluido.

En quinto lugar, se esperaba que las correlaciones entre *Cancelación* y sus puntuaciones de procesamiento (CAA y CAE) fueran altas. Esta predicción se basó en el hecho de que comparten el contenido de los ítems, así como en estudios anteriores (Wechsler, 2003, 2008).

Las tablas A.1-A.9 de este manual muestran las intercorrelaciones de las puntuaciones de las pruebas, las puntuaciones de procesamiento y las puntuaciones compuestas (índices y CI total) para los nueve grupos de edad. Las correlaciones medias de las dos franjas de edad y de los distintos grupos de edad se calcularon utilizando la transformación *z* de Fisher.

La tabla 6.1 presenta las intercorrelaciones medias de las puntuaciones de las pruebas, las puntuaciones de procesamiento y las puntuaciones compuestas para las franjas de edad de 2:6 a 3:11 años y de 4:0 a 7:7 años. Además, incluye las correlaciones de las pruebas con las sumas de las puntuaciones escalares que contribuyen a cada puntuación compuesta. La correlación entre las sumas de las puntuaciones escalares que contribuyen a cada puntuación compuesta y la puntuación escalar de las pruebas se corrigió eliminando la puntuación escalar de cada prueba de la suma de puntuaciones compuestas en la que participaban, como medida de control de los valores inflados. Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Tabla 6.1. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas, por franja de edad

	Edad 2:6-3:11													
Prueba/puntuación de procesamiento/ *puntuación compuesta*	***I***	***D***	***N***	***C***	***RO***	***R***	***L***	***ICV***	***IVE***	***IMT***	***CIT***	***IAV***	***INV***	***ICG***
Información								.44			.56			.54
Dibujos	.44							.44			.46	.42		.41
Nombres	.51	.42										.42		
Cubos	.37	.29	.25						.35		.47		.40	.46
Rompecabezas	.34	.20	.29	.35					.35		.43		.44	.39
Reconocimiento	.42	.39	.37	.34	.37					.48	.54		.53	
Localización	.37	.28	.28	.26	.29	.48				.48			.45	
ICV	.84	.85	.55	.39	.32	.48	.39							
IVE	.44	.30	.33	.83	.82	.43	.34	.43						
IMT	.45	.38	.38	.34	.38	.85	.87	.49	.44					
CIT	.74	.67	.54	.68	.65	.72	.49	.83	.81	.69				
IAV	.56	.84	.84	.32	.29	.45	.34	.83	.37	.45	.71			
INV	.52	.40	.42	.68	.70	.75	.71	.54	.85	.85	.88	.48		
ICG	.76	.69	.52	.72	.67	.54	.43	.85	.84	.55	.97	.71	.82	
Media	10.1	10.0	9.9	9.8	9.9	10.3	10.0	20.1	19.7	20.4	50.1	19.8	40.1	39.8
Dt	3.0	3.1	3.0	3.1	3.0	2.8	3.0	5.1	5.0	5.0	10.4	5.1	8.5	8.6

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Todas las correlaciones entre las pruebas son estadísticamente significativas. El patrón de intercorrelaciones es muy similar al de la WPPSI-III y al de otras escalas de inteligencia de Wechsler (Wechsler, 2002, 2003, 2008).

Edades de 2:6 a 3:11 años

Las correlaciones más altas en esta franja de edad se observan entre las pruebas de Comprensión verbal. Las correlaciones entre las pruebas de Comprensión verbal y las pruebas de Visoespacial son de bajas a moderadas y están dentro de los valores esperados porque todas estas pruebas muestran tener saturaciones altas en *g*. También son moderadas las correlaciones entre las pruebas de Comprensión verbal y las pruebas de Visoespacial. Las pruebas de Comprensión verbal muestran correlaciones más altas con *Reconocimiento* que con *Localización*; lo que puede deberse a que *Reconocimiento* tiene saturaciones más altas en *g*, o bien al tipo de estímulos, dado que, para todas estas pruebas, los estímulos son imágenes. Incluso la prueba *Información*, que incluye ítem que deben responderse verbalmente, empieza con ítems gráficos para los niños pequeños.

El patrón de intercorrelaciones no es tan claro para las pruebas de Visoespacial. La correlación entre *Cubos* y *Rompecabezas* es de las más altas para ambas pruebas; sin embargo, *Cubos* muestra una correlación ligeramente superior con *Información*, y *Rompecabezas* con *Reconocimiento*. Esta ligera diferencia en la magnitud de estas correlaciones probablemente se deba al elevado peso en *g* de estas pruebas, así como a la rendencia de que las aptitudes cognitivas de los niños pequeños estén menos diferenciadas que las de los niños más mayores (Morgan et al., 2009; Tideman y Gustafsson, 2004; Ward et al., 2011). Las pruebas de Visoespacial y de Memoria de trabajo muestran correlaciones de bajas a moderadas. Las correlaciones entre *Reconocimiento* y las pruebas de Visoespacial son más altas que las correlaciones entre *Localización* y las pruebas de Visoespacial. No obstante, podría haberse esperado el patrón inverso tanto porque los requisitos visoespaciales de Localización son superiores a los de *Reconocimiento*, como porque en *Cubos*, *Rompecabezas* y *Localización* se utilizan manipulativos y, por tanto, comparten la varianza del método. El patrón de correlaciones más altas con *Reconocimiento* que con *Localización* se da asimismo con las pruebas de Comprensión verbal, lo que sugiere que *Reconocimiento* tiene una saturación mayor en *g* que *Localización*.

Las pruebas de Memoria de trabajo muestran correlaciones moderadas entre ellas. La correlación entre *Reconocimiento* y *Localización* es la más alta para ambas pruebas.

Aunque las correlaciones varían ligeramente, estos patrones suelen repetirse entre los tres grupos de edad que componen esta franja de edad (véanse las tablas A.1-A.3 de este manual). Estos datos confirman la hipótesis de que la correlación entre pruebas que pretenden medir aptitudes cognitivas similares es mayor que la correlación que obtienen pruebas que miden aptitudes de dominios cognitivos diferentes, lo que supone una demostración inicial de la validez convergente y la validez discriminante de la WPPSI-IV.

Edades de 4:0 a 7:7 años

Las correlaciones más altas en esta franja de edad se observan entre las pruebas de Comprensión verbal. Las correlaciones entre las pruebas de Comprensión verbal y las pruebas de Visoespacial y Razonamiento fluido son moderadas. Estos resultados son los esperados porque, al igual que las pruebas de Comprensión verbal, tanto las pruebas de Visoespacial como las de Razonamiento fluido tienen altas saturaciones en *g*. Las correlaciones entre las pruebas de Comprensión verbal y las pruebas de Memoria de trabajo son de bajas a moderadas. Al igual que en la franja de edad anterior, las pruebas de Comprensión verbal muestran correlaciones más altas con *Reconocimiento* que con *Localización*, lo que sugiere que *Reconocimiento* tiene una saturación mayor en *g* que *Localización*. Las correlaciones entre las pruebas de Comprensión verbal y las pruebas de Velocidad de procesamiento son de bajas a moderadas. En general, las correlaciones entre las pruebas de Comprensión verbal y *Búsqueda de animales* son un poco más altas que las mostradas con *Cancelación* y *Clave de figuras*. Este patrón puede indicar que *Búsqueda de animales* tiene una saturación más alta en *g* que *Cancelación* y *Clave de figuras*.

Las pruebas de Comprensión verbal correlacionan entre sí en mayor medida que con el resto de las pruebas. Las correlaciones entre las pruebas de Visoespacial y las pruebas de Razonamiento fluido son moderadas. Las correlaciones entre *Matrices* y las pruebas de Visoespacial son ligeramente más

altas que las correlaciones entre *Conceptos* y las pruebas de Visoespacial, probablemente debido a que los requisitos de percepción espacial son mayores en *Matrices* que en *Conceptos* (Sattler, 2008a). A diferencia de *Conceptos*, algunos ítems de Matrices son abstractos y requieren que el niño discrimine la posición y la dirección de la figura abstracta, características que comparten con las pruebas de Visoespacial. Las correlaciones entre las pruebas de Visoespacial y las pruebas de Memoria de trabajo son de bajas a moderadas, mostrando las pruebas de Visoespacial correlaciones más altas con *Reconocimiento* que con *Localización*. No obstante, podría haberse esperado el patrón inverso porque los requisitos visoespaciales de *Localización* son superiores a los de *Reconocimiento*. Las correlaciones entre las pruebas de Visoespacial y las pruebas de Velocidad de procesamiento son moderadas. Las correlaciones entre las pruebas de Visoespacial y *Búsqueda de animales* son un poco más altas que las mostradas con *Cancelación* y *Clave de figuras*.

Las correlaciones entre las pruebas de Razonamiento fluido son moderadas. Las correlaciones entre las pruebas de Razonamiento fluido y las pruebas de Comprensión verbal también son moderadas y en algunos casos, igualan la correlación entre *Matrices* y *Conceptos*. Las correlaciones entre las pruebas de Razonamiento fluido y las pruebas de Visoespacial también son moderadas; y, en este caso, la magnitud de la correlación entre *Cubos* y *Matrices*, supera ligeramente la correlación entre *Matrices* y *Conceptos*. Las pruebas de Razonamiento fluido y las pruebas de Memoria de trabajo muestran correlaciones de bajas a moderadas, mostrando las pruebas de Razonamiento fluido correlaciones más altas con *Reconocimiento* que con *Localización*. Las correlaciones entre las pruebas de Razonamiento fluido y las pruebas de Velocidad de procesamiento son de bajas a moderadas. En general, las pruebas de Razonamiento fluido muestran correlaciones un poco más altas con *Búsqueda de animales* que con *Cancelación* o *Clave de figuras*, al igual que sucede con las pruebas de Comprensión verbal y las pruebas de Visoespacial.

Las pruebas de Memoria de trabajo muestran correlaciones moderadas entre ellas. Así como la correlación entre *Reconocimiento* y *Localización* es la más alta para *Localización*, no se da la situación inversa: *Reconocimiento* muestra correlaciones más altas con *Información, Semejanzas, Vocabulario* y *Matrices* que con *Localización*. Las pruebas de Memoria de trabajo y las pruebas de Velocidad de procesamiento muestran correlaciones de bajas a moderadas.

Las pruebas de Velocidad de procesamiento correlacionan entre sí en mayor medida que con el resto de las pruebas; y con las pruebas de Comprensión verbal, las pruebas de Visoespacial y las pruebas de Razonamiento fluido muestran correlaciones más bajas.

Finalmente, las correlaciones entre *Cancelación* y sus correspondientes puntuaciones de procesamiento son muy altas. Este resultado se pronosticó considerando el hecho de que comparten el contenido de los ítems, así como en estudios anteriores (Wechsler, 2003, 2008).

Aunque las correlaciones varían ligeramente, estos patrones suelen repetirse entre los seis grupos de edad que componen esta franja de edad (véanse las tablas A.4-A.9 de este manual). Al igual que se ha señalado para la franja de edad de 2:6 a 3:11 años, estos datos confirman la hipótesis de que la correlación entre pruebas que pretenden medir aptitudes cognitivas similares es mayor que la correlación que obtienen pruebas que miden aptitudes de dominios cognitivos diferentes, lo que supone una demostración inicial de la validez convergente y la validez discriminante de la WPPSI-IV.

Análisis factorial confirmatorio

El objetivo del análisis factorial es identificar un modelo factorial (un conjunto de factores y la estructura de las relaciones entre factores, y entre factores y pruebas) que explique adecuadamente las correlaciones observadas entre las pruebas. Uno de los resultados más importantes del análisis factorial es identificar cuántos factores, y de qué tipo, son necesarios para explicar cómo se interrelacionan las pruebas. Por ejemplo, identificar si en el rendimiento de una prueba influyen tres, cuatro o cinco dimensiones distintas (aptitudes) y el tipo de aptitud cognitiva que se encuentra entre esas dimensiones (p. ej., el razonamiento fluido). De este modo, el análisis factorial relaciona las pruebas de un test con la teoría existente.

Para lograr el objetivo descrito, es útil incluir un gran número de variables en el análisis, puesto que aumenta la validez de constructo y la replicabilidad de los factores, con lo que permite la compa-

ración estadística de modelos alternativos. Por lo tanto, un análisis que incluya varias pruebas de cada tarea y de cada tipo de destreza podrá ofrecer una descripción más completa de los factores subyacentes. Después de haber identificado un modelo factorial óptimo en un grupo de pruebas grande, este mismo modelo puede adecuarse a un grupo de datos reducido (p. ej., las pruebas de los índices primarios) a fin de confirmar que la serie de factores del modelo sigue justificando la matriz de varianza-covarianza.

El modelo factorial de una escala puede evaluarse mediante modelos de ecuaciones estructurales (Jöreskog y Sörbom, 1993). El análisis factorial confirmatorio está diseñado para comprobar una estructura factorial definida previamente por los investigadores a partir de estudios teóricos y empíricos. En este sentido, el análisis factorial confirmatorio es diferente del enfoque exploratorio, porque la agrupación de las pruebas se realiza previamente y no procede de un algoritmo estadístico. Las relaciones específicas entre las variables latentes (p. ej., la aptitud de comprensión verbal) y entre las variables latentes y las variables observadas (en este caso, las pruebas), se especifican previamente en un modelo; y este modelo se pone a prueba para determinar si ofrece una buena justificación de las covarianzas entre las variables observadas.

El análisis factorial confirmatorio es preferible al análisis factorial exploratorio cuando se dispone de una teoría explícita de la estructura factorial o cuando se han desarrollado modelos alternativos al respecto (Schmitt, 2011; Stevens, 1996). A lo largo de 70 años se han realizado numerosas investigaciones con las escalas de inteligencia de Wechsler. Durante este tiempo, los avances en la teoría de la inteligencia y las mejoras logradas en las técnicas psicométricas han guiado las revisiones de las escalas Wechsler. Los estudios posteriores han sometido la WPPSI, la *Wechsler Intelligence Scale for Children* (WISC; Wechsler, 1949) y la *Wechsler Adult Intelligence Scale* (WAIS; Wechsler, 1955), así como sus respectivas sucesoras, a evaluaciones adicionales de las estructuras de ambas escalas. Tal y como demostraron los resultados del análisis factorial, la incorporación de pruebas nuevas (p. ej., *Claves* y *Búsqueda de símbolos*) proporcionaron medidas fiables y distintivas de nuevos constructos de aptitud (p. ej., Velocidad de procesamiento).

Durante años, los estudios de análisis factorial han indicado que la WISC-IV y la WAIS-IV (y sus predecesoras) miden cuatro dominios cognitivos (p. ej., Bodin et al. 2009; Bowden, Weiss, Holdnack y Lloyd, 2006; Ward et al., 2011; Watkins, 2010; Watkins et al., 2006; Wechsler, 2003, 2008). Además, la estructura de cuatro factores ha demostrado ser muy robusta en las adaptaciones realizadas en otros países (p. ej., Chen, Keith, Weiss, Zhu, y Li, 2010). Algunos autores, basándose en otras perspectivas y teorías, evaluaron la adecuación de modelos de dos o tres factores y concluyeron que dichos modelos podían igualmente explicar adecuadamente los datos (Kaufman, Lichtenberger y McLean, 2001; Ward, Ryan y Axelrod, 2000). Como ya se ha mencionado anteriormente, los investigadores han identificado una estructura de cinco factores alternativa y común a la WISC-IV y la WAIS-IV, que puede tener ventajas (Benson et al., 2010; Keith et al., 2006; Weiss et al., 2013; Weiss et al., 2013). Los nombres de los factores varían en función del sistema taxonómico seguido por el equipo investigador y vienen a coincidir con Comprensión verbal, Visoespacial, Razonamiento fluido, Memoria de trabajo y Velocidad de procesamiento.

El desarrollo de la WPPSI-IV se basó en el supuesto teórico de que la escala ofrece una estimación de la aptitud cognitiva general (el CI total), medida por tres dominios cognitivos para los niños de 2:6 a 3:11 años (comprensión verbal, visoespacial y memoria de trabajo) y por cinco dominios cognitivos para los niños de 4:0 a 7:7 años (comprensión verbal, visoespacial razonamiento fluido, memoria de trabajo y velocidad de procesamiento). Las pruebas *Reconocimiento* y *Localización* se diseñaron para medir la memoria de trabajo; y *Búsqueda de animales, Cancelación* y *Clave de figuras* se añadieron como medidas de la velocidad de procesamiento. El análisis factorial exploratorio se utilizó en las primeras fases de desarrollo de las pruebas para evaluar el patrón de relaciones entre las pruebas nuevas y las conservadas respecto a la WPPSI-III.

La estructura de la escala, descrita en el capítulo 2 de este manual, se basa en los resultados obtenidos en las investigaciones que se efectuaron en las primeras fases, y propone un modelo factorial para cada una de las franjas de edad de la WPPSI-IV. Para la franja de edad de 2:6 a 3:11 años, se propone un modelo factorial de segundo orden, con tres factores de primer orden (Comprensión verbal, Visoespacial y Memoria de trabajo), que contribuyen a un factor *g* de segundo orden (el factor Escala total). Para la franja de edad de 4:0 a 7:7 años, se propone un modelo factorial

también de segundo orden, con cinco factores de primer orden (Comprensión verbal, Visoespacial, Razonamiento fluido, Memoria de trabajo y Velocidad de procesamiento), que contribuyen a un factor *g* de segundo orden (el factor Escala total). Para evaluar la estructura factorial de la WPPSI-IV, se realizaron análisis factoriales confirmatorios. En particular, los análisis tenían como objetivo esclarecer el número y tipo de factores identificados por el instrumento, y determinar hasta qué punto las pruebas de la WPPSI-IV miden los constructos de aptitud latente propuestos.

La estrategia de investigación de la estructura factorial de la WPPSI-IV consistió, primero, en el análisis de todas las pruebas de cada franja de edad (2:6-3:11 y 4:0-7:7). Tras identificar un modelo factorial óptimo para el conjunto de todas las pruebas de la WPPSI-IV en la franja de edad de 4:0-7:7 años, se evaluó el ajuste de dicho modelo en los tres grupos de edad siguientes: 4:0-4:11, 5:0-5:11 y 6:0–7:7. Posteriormente se ajustaron los modelos óptimos a las pruebas de los índices primarios para confirmar que los factores del modelo seguían explicando la matriz de varianza-covarianza.

Análisis factorial confirmatorio de todas las pruebas

La mayoría de las medidas de bondad de ajuste utilizadas para evaluar modelos factoriales se basan en el estadístico chi cuadrado. La prueba χ^2 (chi cuadrado) se utiliza para examinar la hipótesis nula según la cual la matriz de varianza-covarianza basada en las cargas factoriales y las intercorrelaciones factoriales es igual que la matriz de varianza-covarianza observada (Byrne, 2001; Schumacker y Lomax, 2004; Thompson, 2000). Por lo tanto, si el modelo se ajusta adecuadamente a los datos, la hipótesis nula no se rechaza.

Debido a que el valor de χ^2 (chi cuadrado) tiende a ser más sensible en muestras de gran tamaño, se recomienda utilizar otras pruebas estadísticas que no sean dependientes del tamaño de la muestra, puesto que se pueden producir estimaciones sesgadas del ajuste del modelo (Byrne, 2001; Schumacker y Lomax, 2004; Thompson, 2000). No obstante, la prueba χ^2 sigue siendo necesaria dada su utilidad para examinar la significación estadística de la diferencia en el ajuste de dos modelos.

Para determinar el grado en que la matriz de varianza-covarianza observada en la muestra de tipificación y la estimada son equivalentes, generalmente se usan seis índices adicionales de bondad de ajuste. El *AGFI* (índice ajustado de bondad de ajuste; Jöreskog y Sörbom, 1993) es la proporción de la matriz de la varianza-covarianza observada, explicada por la matriz de la varianza-covarianza estimada, y ajustada según los grados de libertad utilizados en el modelo (Schumacker y Lomax, 2004). Los valores para este estadístico deberían ser superiores a .90 cuando el tamaño de la muestra es más grande, o .95 cuando el tamaño de la muestra es más pequeño (Hooper, Coughlan y Mullen, 2008; Thompson, 2000). El *CFI* (índice de ajuste comparativo) examina cómo el modelo específico reproduce la matriz de la varianza-covarianza observada en comparación con un modelo independiente. Hu y Bentler (1999) recomiendan un valor de *CFI* de .95 o superior. El *TLI* (índice de Tucker-Lewis) mide el ajuste incremental de una forma similar al *CFI*, con la diferencia de estar corregido por los grados de libertad del modelo. Para analizar el ajuste del modelo respecto a la matriz de covarianza de la población, se ha utilizado el error de aproximación cuadrático medio (*RMSEA*) de Steiger (1990). El *RMSEA* es una estimación de cómo la varianza-covarianza estimada reproduce la matriz de covarianza de la población, corregida por los grados de libertad del modelo ajustado. Browne y Cudeck (1993) sugieren que un valor de RMSEA de .05 o menor indica un ajuste próximo al modelo y que los valores inferiores a .08 representan un ajuste adecuado con errores de aproximación razonables en la población. Asimismo se han incluido el *AIC* (criterio de información de Akaike; Akaike, 1974) y el *BIC* (criterio de información bayesiano; Schwarz, 1978). Cuando se comparan dos modelos con estos estadísticos, los valores más bajos representan un ajuste mejor.

Se examinó la bondad de ajuste de diversos modelos para las dos franjas de edad de la WPPSI-IV: de 2:6 a 3:11 años y de 4:0 a 7:7 años. Excepto que se indique lo contrario, cada modelo tiene dos o más factores de primer orden y un único factor *g* de segundo orden. El modelo seleccionado para representar la estructura de la WPPSI-IV correspondiente a la franja de edad de 4:0 a 7:7 años también se probó en tres grupos de edad: 4:0-4:11, 5:0-5:11 y 6:0–7:7. En los apartados siguientes se describen los modelos para cada franja de edad, así como sus resultados.

Modelos para las edades de 2:6 a 3:11 años

En el modelo 1, todas las pruebas (siete) contribuyen al factor *g*. El modelo 2 representa la estructura tradicional de dos factores, con el factor verbal y el factor manipulativo separados. El modelo 3 representa la estructura de tres factores, con los factores Comprensión verbal, Visoespacial y Memoria de trabajo. A continuación se presenta la estructura de las pruebas que componen cada modelo:

- Modelo 1 (un factor de primer orden)

  7 pruebas en un único factor general

- Modelo 2 (un factor de segundo orden y dos factores de primer orden)

  3 pruebas de Comprensión verbal en el primer factor.

  2 pruebas de Visoespacial y 2 pruebas de Memoria de trabajo en el segundo factor.

- Modelo 3 (un factor de segundo orden y tres factores de primer orden)

  3 pruebas de Comprensión verbal en el primer factor.

  2 pruebas de Visoespacial en el segundo factor.

  2 pruebas de Memoria de trabajo en el tercer factor.

Modelos para las edades de 4:0 a 7:7 años

En el modelo 1, todas las pruebas (quince) contribuyen al factor *g*. El modelo 2 representa la estructura tradicional de dos factores, con el factor verbal y el factor manipulativo separados. El modelo 3 permite examinar la distinción de las pruebas de Velocidad de procesamiento respecto a las tradicionalmente denominadas *pruebas manipulativas* (pruebas de Visoespacial, de Razonamiento fluido y de Memoria de trabajo). Es un modelo de tres factores con un factor Comprensión verbal; un factor combinado Visoespacial, Razonamiento fluido y Memoria de trabajo; y un factor Velocidad de procesamiento.

Los modelos 4a y 4b son dos variantes de los modelos de cuatro factores. El modelo 4a permite validar la distinción de las pruebas de Memoria de trabajo respecto a las pruebas de Visoespacial y de Razonamiento fluido y refleja los modelos publicados de la WISC-IV y la WAIS-IV (un modelo de cuatro factores con un factor Comprensión verbal, un factor combinado Visoespacial y Razonamiento fluido, un factor Memoria de trabajo y un factor Velocidad de procesamiento). El modelo 4b se basa en los estudios de neurociencia cognitiva que indican que el razonamiento fluido y la memoria de trabajo comparten una función del córtex prefrontal. También es un modelo de cuatro factores, con un factor Comprensión verbal; un factor Visoespacial, un factor combinado Razonamiento fluido y Memoria de trabajo, y un factor Velocidad de procesamiento.

Los modelos 5a y 5b son variantes de un modelo de cinco factores. El modelo 5a incluye un factor Comprensión verbal; un factor Visoespacial, un factor Razonamiento fluido, un factor Memoria de trabajo y un factor Velocidad de procesamiento. El modelo 5 b es igual que el modelo 5a, pero divide el factor Comprensión verbal en dos subfactores. A continuación se presenta la estructura de las pruebas que componen cada modelo:

- Modelo 1 (un factor de primer orden)

  15 pruebas en un único factor general.

- Modelo 2 (un factor de segundo orden y dos factores de primer orden)

  6 pruebas de Comprensión verbal en el primer factor.

  2 pruebas de Visoespacial, 2 pruebas de Razonamiento fluido, 2 pruebas de Memoria de trabajo y 3 pruebas de Velocidad de procesamiento en el segundo factor.

- Modelo 3 (un factor de segundo orden y tres factores de primer orden)

  6 pruebas de Comprensión verbal en el primer factor.

  2 pruebas de Visoespacial, 2 pruebas de Razonamiento fluido y 2 pruebas de Memoria de trabajo en el segundo factor.

  3 pruebas de Velocidad de procesamiento en el tercer factor.

- Modelo 4a (un factor de segundo orden y cuatro factores de primer orden)
  - 6 pruebas de Comprensión verbal en el primer factor.
  - 2 pruebas de Visoespacial y 2 pruebas de Razonamiento fluido en el segundo factor.
  - 2 pruebas de Memoria de trabajo en el tercer factor.
  - 3 pruebas de Velocidad de procesamiento en el cuarto factor.
- Modelo 4b (un factor de segundo orden y cuatro factores de primer orden)
  - 6 pruebas de Comprensión verbal en el primer factor.
  - 2 pruebas de Visoespacial en el segundo factor.
  - 2 pruebas de Razonamiento fluido y 2 pruebas de Memoria de trabajo en el tercer factor.
  - 3 pruebas de Velocidad de procesamiento en el cuarto factor.
- Modelo 5a (un factor de segundo orden y cinco factores de primer orden)
  - 6 pruebas de Comprensión verbal en el primer factor.
  - 2 pruebas de Visoespacial en el segundo factor.
  - 2 pruebas de Razonamiento fluido en el tercer factor.
  - 2 pruebas de Memoria de trabajo en el cuarto factor.
  - 3 pruebas de Velocidad de procesamiento en el quinto factor.
- Modelo 5b (un factor de segundo orden, cinco factores de primer orden y dos subfactores en el factor Comprensión verbal)
  - 6 pruebas de Comprensión verbal en el primer factor.
    - *Información, Semejanzas, Vocabulario* y *Comprensión* en el primer subfactor.
    - *Dibujos* y *Nombres* en el segundo subfactor.
  - 2 pruebas de Visoespacial en el segundo factor.
  - 2 pruebas de Razonamiento fluido en el tercer factor.
  - 2 pruebas de Memoria de trabajo en el cuarto factor.
  - 3 pruebas de Velocidad de procesamiento en el quinto factor.

Resultados

La tabla 6.2 muestra los resultados del análisis factorial confirmatorio de todas las pruebas tanto para las dos franjas de edad como para los tres grupos de edad siguientes: 4:0-4:11, 5:0-5:11 y 6:0–7:7.

Tabla 6.2. Índices de bondad de ajuste para el análisis factorial confirmatorio de todas las pruebas

	Índice de bondad de ajuste								Incremento		
Modelo	χ^2	*gl*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*AIC*	*BIC*	$\Delta\chi^2$	Δgl	*p*
Edad 2:6-3:11 (*n* = 288)											
Modelo independiente	339.86	21									
Modelo 1	36.98	14	0.93	0.93	0.89	0.08	64.98	116.21	302.89	7	<.001
Modelo 2	16.81	10	0.95	0.98	0.96	0.05	52.81	118.68	20.17	4	<.001
Modelo 3	9.06	8	0.97	1.00	0.99	0.02	49.06	122.25	7.75	2	0.021
Edad 4:0-7:7 (*n* = 578)											
Modelo independiente	2188.78	105									
Modelo 1	481.18	90	0.84	0.81	0.78	0.09	541.18	671.81	1707.60	15	<.001
Modelo 2	325.37	86	0.89	0.89	0.86	0.07	393.37	541.42	155.81	4	<.001
Modelo 3	191.39	84	0.94	0.95	0.94	0.05	263.39	420.14	133.98	2	<.001
Modelo 4a	185.12	82	0.94	0.95	0.94	0.05	261.12	426.58	6.27	2	0.044
Modelo 4b[a]	180.8	82	0.94	0.95	0.94	0.05	256.80	422.27	10.58	2	0.005
Modelo 5a[b]	177.67	80	0.94	0.95	0.94	0.05	257.67	431.84	3.14	2	0.208
Modelo 5b[c]	159.41	76	0.94	0.96	0.94	0.04	247.41	439.00	18.26	4	0.001
Edad 4:0-4:11 (*n* = 199)											
Modelo 5b	117.7	76	0.89	0.95	0.93	0.05	205.70	350.17			
Edad 5:0-5:11 (*n* = 195)											
Modelo 5b	104.24	76	0.90	0.97	0.95	0.04	192.24	336.02			
Edad 6:0-7:7 (*n* = 184)											
Modelo 5b	118.35	76	0.88	0.92	0.89	0.06	206.35	347.81			

Nota: Los valores chi cuadrado están calculados con el método de mínimos cuadrados ponderados (WLS) del paquete estadístico 9.3 de SAS.

[a]La mejora del modelo se basa en la comparación con el modelo 3.

[b]La mejora del modelo se basa en la comparación con el modelo 4b.

[c]La mejora del modelo se basa en la comparación con el modelo 5a.

Los resultados que figuran en la tabla 6.2 muestran que el modelo de tres factores (modelo 3) se ajusta mejor a la franja de edad de 2:6 a 3:11 años. La comparación de los índices de ajuste de los modelos 1 y 2 indica que hay una mejora estadísticamente significativa del ajuste para el modelo 2. La comparación de los índices de ajuste de los modelos 2 y 3 indica que hay una mejora estadísticamente significativa para el modelo 3. Para la franja de edad de 2:6 a 3:11 años, las intercorrelaciones factoriales de una versión no jerárquica del modelo 3 van de .68 (entre Comprensión verbal y Memoria de trabajo) a .71 (entre Visoespacial y Memoria de trabajo).

El análisis del modelo 3 sugirió que dos modificaciones mejorarían el ajuste total, concretamente, haciendo que las pruebas *Información* y *Nombres* y las pruebas *Dibujos* y *Reconocimiento* tuvieran errores correlacionados. Permitir el error correlacionado es una forma de tener en cuenta una característica común de dos pruebas que hace que las correlaciones entre ellas sean más altas (o menos altas) que lo que se podría esperar por el modelo factorial. En el caso de *Información* y *Nombres*, esta característica común podría ser que son las únicas pruebas, de entre las que componen la batería para los niños más pequeños, que deben responderse verbalmente. En el caso de *Dibujos* y *Reconocimiento*, la característica común podría ser que requieren la identificación de una imagen. No obstante, se decidió no incorporar estas modificaciones al modelo factorial de la WPPSI-IV. Aunque el ajuste habría mejorado en cada uno de estos casos, el hecho de permitir los errores correlacionados no mejoraría la interpretación del modelo. Por ello, se seleccionó el

modelo 3 para representar la estructura de la WPPSI-IV correspondiente a todas las pruebas de la franja de edad de 2:6 a 3:11 años. En la figura 6.1 se ilustra este modelo, con las cargas factoriales.

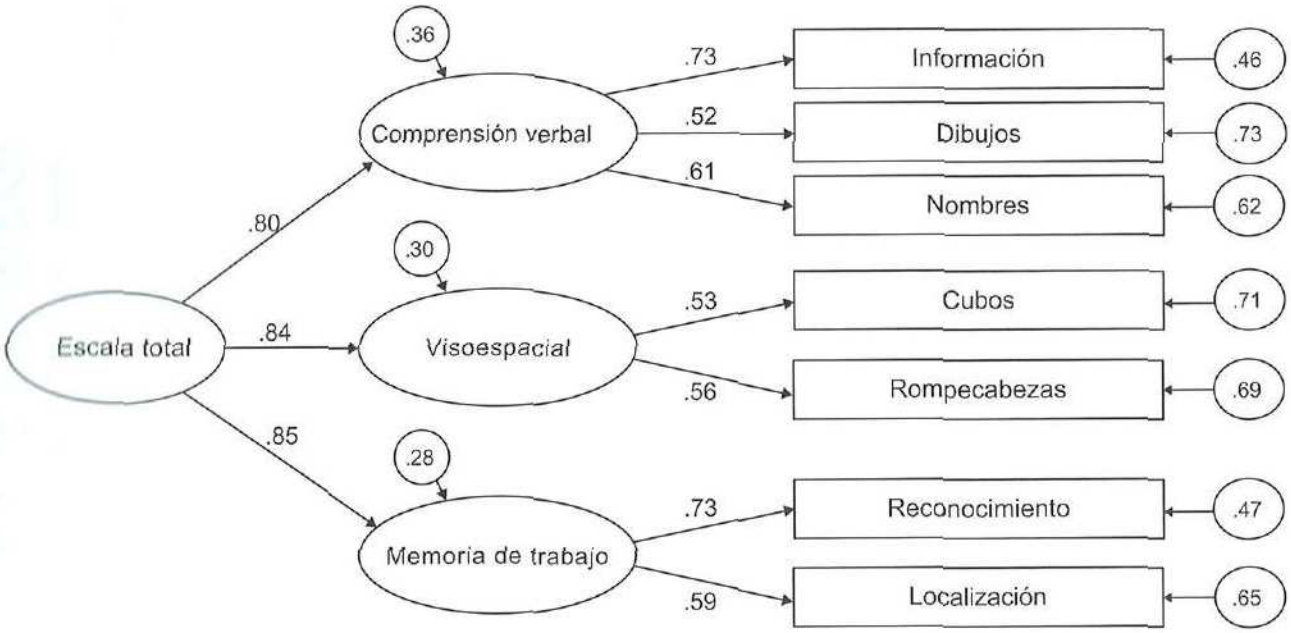


Figura 6.1. Modelo jerárquico de tres factores para todas las pruebas (2:6-3:11 años)

Los resultados que figuran en la tabla 6.2 muestran que los modelos de cinco factores (modelos 5a y 5b) se ajustan mejor a la franja de edad de 4:0 a 7:7 años. Los modelos de uno y dos factores no ofrecieron un ajuste adecuado, mientras que el resto de los modelos ofrecieron ajustes adecuados. Al separar el factor Velocidad de procesamiento del factor combinado Visoespacial, Razonamiento fluido y Memoria de trabajo (modelo 3) se obtuvo una mejora estadísticamente significativa del ajuste del modelo. Del mismo modo, la comparación de los índices de los modelos 3 y 4a, en los que el factor Memoria de trabajo se había separado del factor combinado Visoespacial y Razonamiento fluido, indicó una mejora moderada en el ajuste, pero estadísticamente significativa. Una mejora similar se obtuvo al comparar los índices de ajuste de los modelos 3 y 4b, en los que el factor Visoespacial se había separado del factor combinado Razonamiento fluido y Memoria de trabajo.

El *modelo 5a, de cinco factores* (Comprensión verbal, Visoespacial, Razonamiento fluido, Memoria de trabajo y Velocidad de procesamiento), obtuvo un ajuste significativamente mejor en comparación con los dos modelos de cuatro factores (4a y 4b). Los valores de la mejora en el ajuste entre los modelos 4a y 5a ($\Delta\chi^2/\Delta$ *df* = 7.45, 2, *p* = .024) no figuran en la tabla porque el modelo 4b se ajusta a los datos mejor que el modelo 4a. El ajuste del modelo 5a fue excelente; no obstante, los índices de modificación y los residuales sugirieron la presencia de *dos subfactores dentro del factor Comprensión verbal*. El *primer subfactor, compuesto* por *Información*, *Semejanzas*, *Vocabulario* y *Comprensión*, se denominó *Amplio/Expresivo*; y el segundo factor, compuesto por *Dibujos* y *Nombres*, se denominó *Restringido/Simple*. Estas subagrupaciones son significativas teóricamente. Las pruebas del subfactor Amplio/Expresivo requieren más expresión o, en el caso de *Información* y *Comprensión*, no se limitan al conocimiento de los significados de las palabras. El subfactor Restringido/Simple está formado por dos medidas del vocabulario que no requieren expresión o requieren poca. El *modelo 5b es*, por tanto, una modificación del modelo 5a e incluye estos dos subfactores. La comparación de los índices de ajuste de los modelos 5a y 5b indica que el ajuste del modelo 5b es ligeramente mejor que el del modelo 5a, y la mejora en el ajuste es estadísticamente significativa. El modelo 5b también produjo un buen ajuste para cada uno de los tres grupos de la franja de más edad: 4:0-4:11, 5:0-5:11 y 6:0–7:7. Por ello, se seleccionó el modelo 5b para representar la estructura de la WPPSI-IV correspondiente a todas las pruebas de la franja de edad de 4:0 a 7:7 años. En la figura 6.2 se ilustra este modelo, con las cargas factoriales.

Además, se examinó una versión no jerárquica del modelo 5a para evaluar las intercorrelaciones de los cinco factores de este modelo. (Dado que el modelo 5b difiere del modelo 5a únicamente en la división del factor Comprensión verbal en dos subfactores, las intercorrelaciones de los cinco factores del modelo 5b son muy similares a los del modelo 5a). Las intercorrelaciones factoriales van de .63 (entre Comprensión verbal y Velocidad de procesamiento) a .88 (entre Visoespacial y Razonamiento fluido).

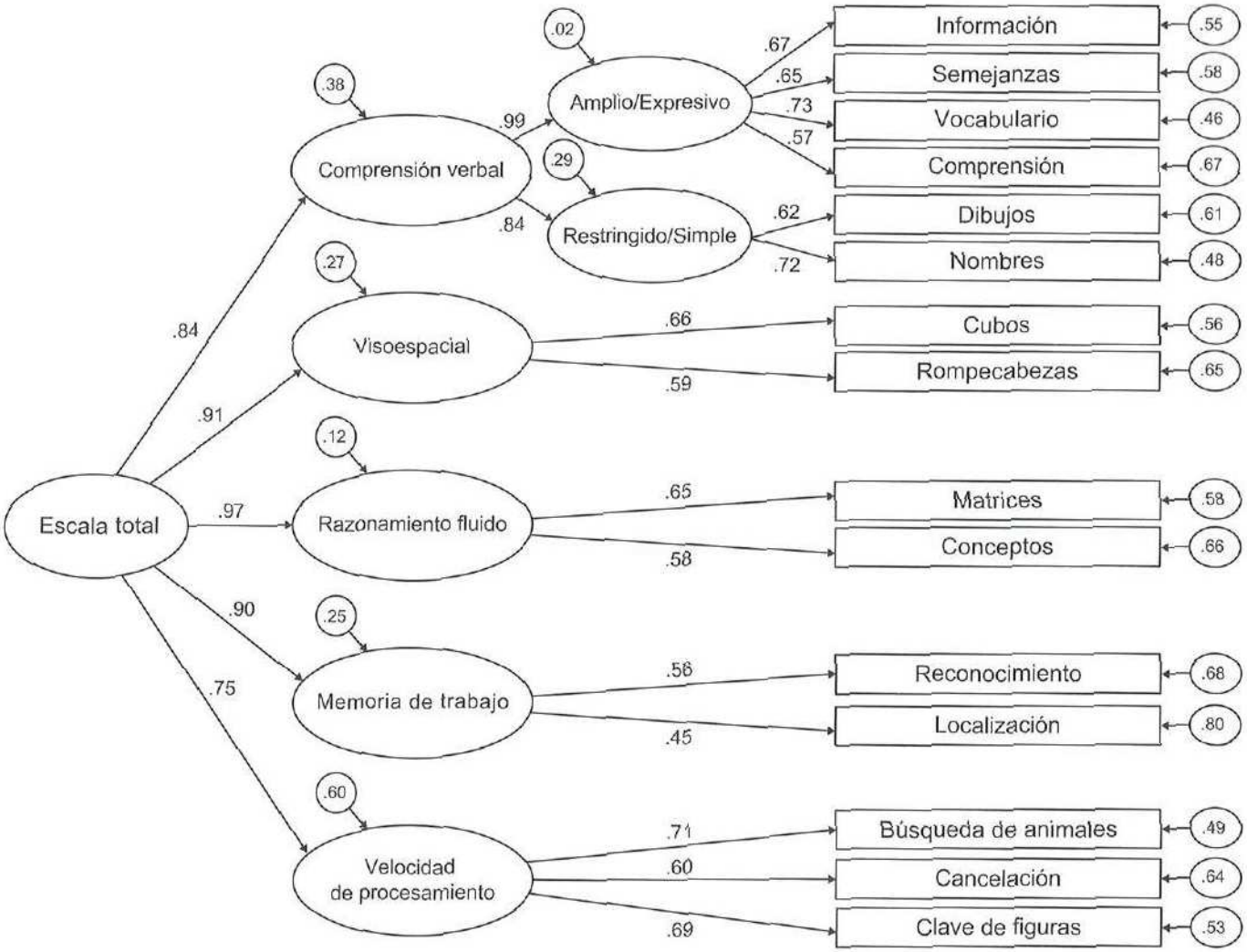


Figura 6.2. Modelo jerárquico de cinco factores para todas las pruebas (4:0-7:7 años)

Análisis de las pruebas de los índices primarios

Los modelos seleccionados para representar la estructura de la WPPSI-IV se ajustaron a las pruebas que conforman los índices primarios para las dos franjas de edad (2:6-3:11 y 4:0-7:7). En esta fase, el objetivo de la investigación no se centró en el número de factores que influyen en las pruebas, sino en cuán bien los factores identificados justifican las intercorrelaciones del reducido grupo de pruebas que conforman los índices primarios. En los apartados siguientes se describen los modelos para cada franja de edad, así como los resultados.

Modelos para las edades de 2:6 a 3:11 años

En la primera fase del análisis (cuando se incluían todas las pruebas) se seleccionó el modelo 3, de tres factores. En este nuevo análisis, en vez de siete pruebas, solo se incluyeron las seis pruebas que conforman los índices primarios; se omitió *Nombres* porque no contribuye a ningún índice primario.

A continuación se presentan las pruebas que componen el modelo:

- Modelo 3 (un factor de segundo orden y tres factores de primer orden)
  - 2 pruebas de Comprensión verbal en el primer factor.
  - 2 pruebas de Visoespacial en el segundo factor.
  - 2 pruebas de Memoria de trabajo en el tercer factor.

Modelos para las edades de 4:0 a 7:7 años

Cuando todas las pruebas se incluían en el análisis, se seleccionó el modelo 5b, de cinco factores. Sin embargo, dado que *Dibujos* y *Nombres* no contribuyen a ningún índice primario, en este análisis no se incluyeron los subfactores del factor Comprensión verbal. De este modo, cuando en el análisis se incluyen solo las diez pruebas que conforman los índices primarios, el modelo apropiado es el 5a. A continuación se presentan las pruebas que componen el modelo:

- Modelo 5a (un factor de segundo orden y cinco factores de primer orden)
  - 2 pruebas de Comprensión verbal en el primer factor.
  - 2 pruebas de Visoespacial en el segundo factor.
  - 2 pruebas de Razonamiento fluido en el tercer factor.
  - 2 pruebas de Memoria de trabajo en el cuarto factor.
  - 2 pruebas de Velocidad de procesamiento en el quinto factor.

Resultados

La tabla 6.3 muestra los resultados del análisis factorial confirmatorio de las pruebas de los índices primarios tanto para las dos franjas de edad.

Tabla 6.3. Índices de bondad de ajuste para el análisis factorial confirmatorio de las pruebas de los índices primarios

	Índice de bondad de ajuste							
Modelo	χ^2	***gl***	***AGFI***	***CFI***	***TLI***	***RMSEA***	***AIC***	***BIC***
Edad 2:6-3:11 (*n* = 288)								
Modelo 3	6.34	3	0.95	0.99	0.93	0.06	42.34	108.21
Edad 4:0-7:7 (*n* = 578)								
Modelo 5a	38.39	25	0.97	0.99	0.98	0.03	98.39	229.07

Nota: Los valores chi cuadrado están calculados con el método de mínimos cuadrados ponderados (WLS) del paquete estadístico 9.3 de SAS.

Los resultados que figuran en la tabla 6.3 muestran que el modelo de tres factores (modelo 3) ofrece un buen ajuste para las pruebas de los índices primarios correspondientes a la franja de edad de 2:6 a 3:11 años.

Los resultados presentados en la tabla 6.3 indican que el modelo de cinco factores (modelo 5a) ofrece un buen ajuste para las pruebas de los índices primarios correspondientes a la franja de más edad. Las figuras 6.3 y 6.4 ilustran la estructura de la WPPSI-IV para cada franja de edad, utilizando solo las pruebas que conforman los índices primarios.

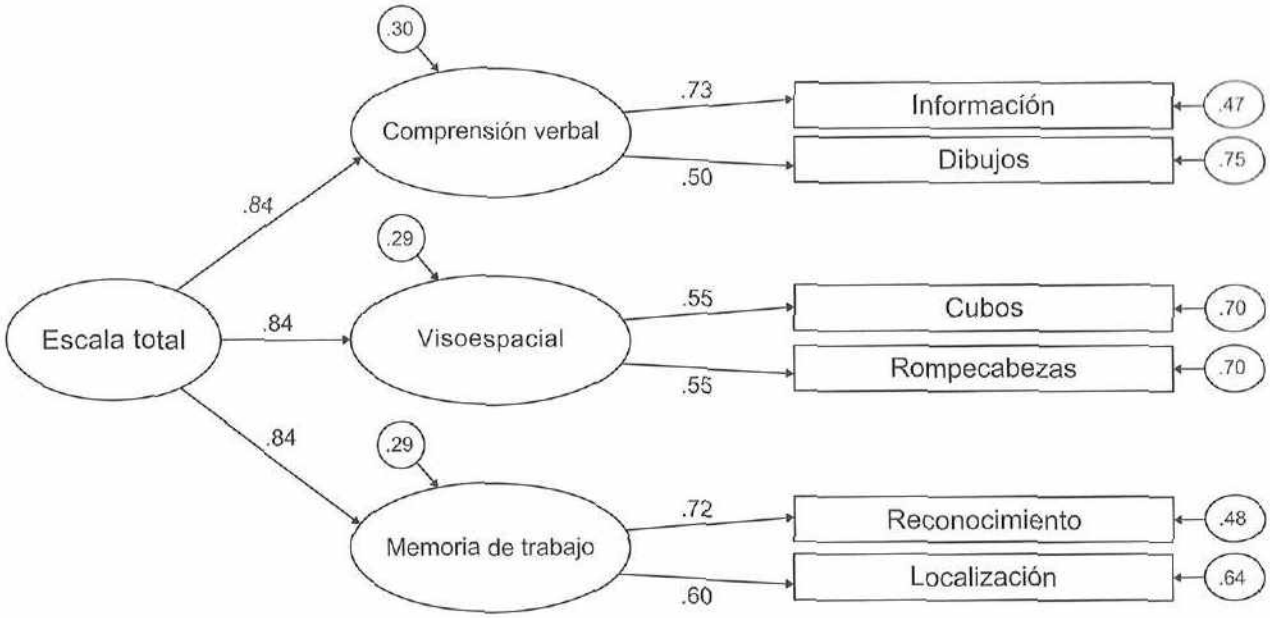


Figura 6.3. Modelo jerárquico de tres factores para las pruebas de los índices primarios (2:6-3:11 años)

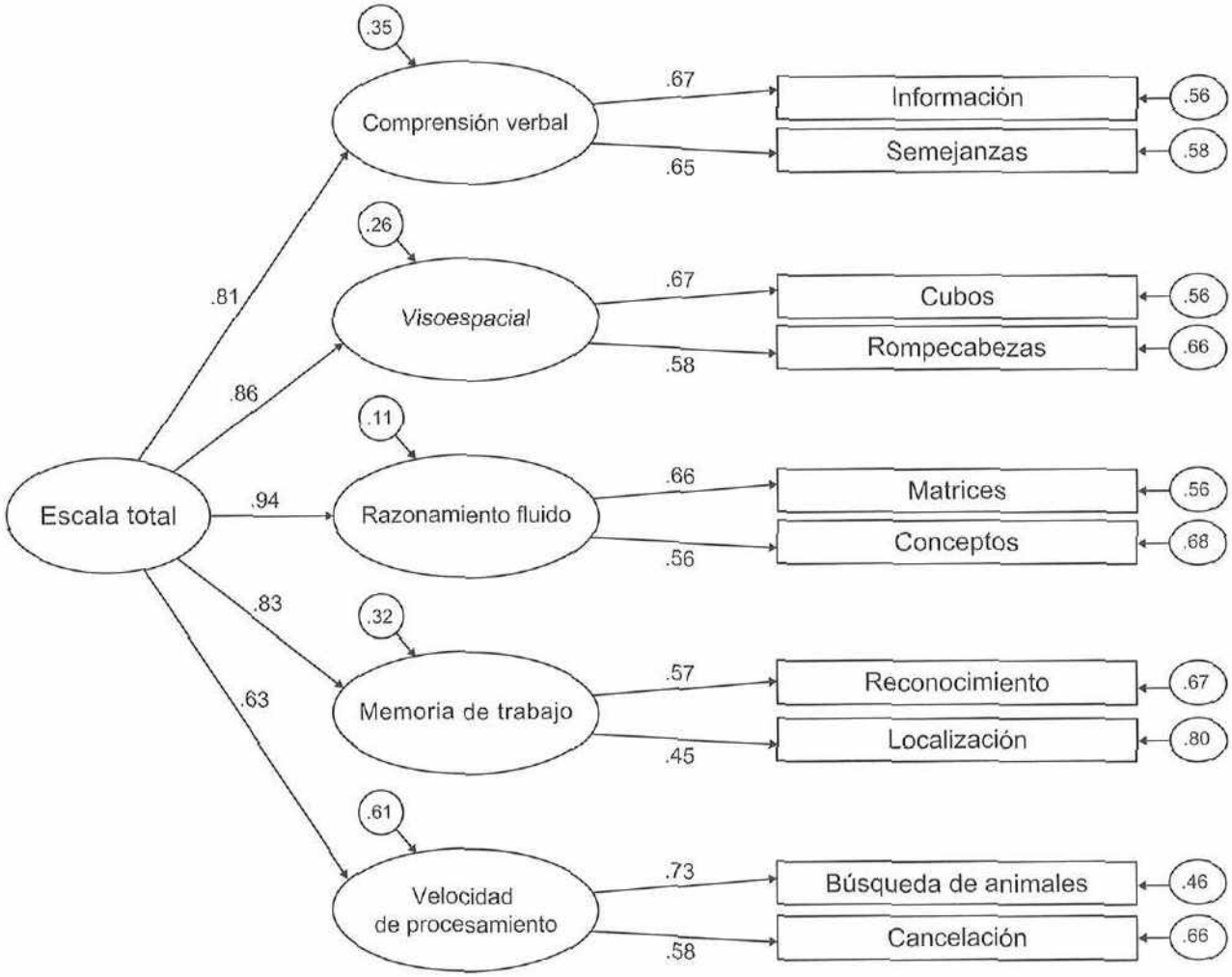


Figura 6.4. Modelo jerárquico de cinco factores para las pruebas de los índices primarios (4:0-7:7 años)

Evidencias basadas en las relaciones con otras variables

El estudio de la relación entre las puntuaciones de los tests con variables externas relacionadas proporciona una evidencia adicional de su validez. A menudo esta evidencia se obtiene examinando las relaciones de un test con otros instrumentos que han sido diseñados para medir el mismo constructo o un constructo similar. Durante el desarrollo de la WPPSI-IV se seleccionaron, por su valor como instrumentos de evaluación, otras medidas de aptitud cognitiva o intelectual, así como medidas del funcionamiento ejecutivo, del funcionamiento conductual y emocional. Históricamente, este tipo de evidencia ha recibido diversas denominaciones: *validez concurrente, validez convergente y discriminante, validez predictiva* y *validez de criterio*. Pero, independientemente de la terminología, el análisis de las relaciones entre las puntuaciones de un test y otras variables similares conceptualmente proporciona una información muy valiosa respecto a lo que mide el test y si cumple las expectativas. Los datos relativos a la validez de la WPPSI-IV que se describen en esta sección se dividen en dos apartados. En el primero se presentan una serie de estudios llevados a cabo durante el proceso de tipificación estadounidense y español para analizar las relaciones entre las puntuaciones de la WPPSI-IV y otras medidas de aptitud y rendimiento intelectual. En el segundo se describen los estudios estadounidenses realizados con grupos especiales que respaldan la validez de la WPPSI-IV para distintos grupos clínicos de niños (p. ej., alta capacidad intelectual, discapacidad intelectual, TDAH, trastorno autista).

Relaciones con otras medidas

Cronbach y Meehl (1955) consideran que las evidencias de validez obtenidas a partir de las relaciones con otras medidas permiten situar los constructos medidos dentro de una red nomológica de variables conocidas. Campbell y Fiske (1959) también abordaron la importancia de establecer asociaciones entre constructos estrechamente relacionados (validez convergente) y entre constructos vagamente relacionados (validez discriminante).

Con las muestras no clínicas se analizaron las relaciones entre la WPPSI-IV y los siguientes instrumentos de evaluación: WPPSI-III, Bayley-III, WISC-IV, NNAT2, NEPSY-II y BASC–2 CPP. La tabla 6.4 recoge el tamaño de la muestra y los datos demográficos para cada uno de los estudios comparativos. La WPPSI-IV y la mayoría de las otras medidas de aptitud intelectual (WPPSI-III, Bayley-III y WISC-IV) se aplicaron en orden contrabalanceado. En los estudios contrabalanceados, las medias, las desviaciones típicas y los coeficientes de correlación corregidos se calcularon por separado para cada parte de la muestra, según el orden de aplicación y, posteriormente, se calculó el promedio. Este método evita que los efectos de aplicación repetida reduzcan artificialmente los coeficientes de correlación. En el estudio de correlación de la WPPSI-IV con la NNAT2 se aplicó la WPPSI-IV en primer lugar.

Tabla 6.4. Datos demográficos para los estudios de validez con otras medidas

	WPPSI-III*	Bayley-III	WISC-IV	NNAT2	NEPSY-II	BASC-2 CPP
N	49	115	99	62	55	1984
Edad						
Media	4.8	3.0	6.8	4.5	5.2	4.9
Dt	1.3	0.3	0.4	0.3	1.3	1.4
Sexo						
Mujer	51.0	51.3	45.5	58.1	34.5	51.7
Varón	49.0	48.7	54.5	41.9	65.5	48.3

Nota: Excepto para el tamaño de la muestra (*N*) y la edad, los datos se muestran en porcentajes. Si el porcentaje total no suma 100, es debido al redondeo.

* Datos españoles.

Correlaciones con la WPPSI-III

Se aplicó la WPPSI-IV y la WPPSI-III a 49 niños de 2:9 a 7:5 años, en orden contrabalanceado, con un intervalo de aplicación de 13 a 69 días (media = 29 días). La tabla 6.5 muestra las medias, las desviaciones típicas, las correlaciones (corregidas y sin corregir) y las diferencias típicas.

Tabla 6.5. Correlaciones entre la WPPSI-IV y la WPPSI-III (datos españoles)

Prueba/puntuación de procesamiento/ puntuación compuesta	WPPSI-IV			WPPSI-III					
	Media[a]	*Dt*[a]	*n*	Media[a]	*Dt*[a]	*n*	r_{12}[b]	r_{12}[b] corregido	Diferencia típica[c]
Información	10.9	2.2	48	11.1	2.1	48	.61	.72	.08
Semejanzas	9.8	1.8	22	9.6	1.9	22	.77	.90	.12
Vocabulario	9.4	2.3	22	9.3	2.3	22	.54	.64	.04
Comprensión	9.5	2.8	22	8.5	2.8	22	.66	.69	.36
Dibujos	10.6	2.7	47	12.2	2.8	47	.49	.53	.57
Nombres	10.2	2.7	49	10.9	2.7	49	.73	.76	.28
Adivinanzas				10.8	2.8	24			
Cubos	9.9	2.5	47	9.7	2.3	47	.76	.81	.12
Rompecabezas	9.9	2.4	48	9.7	2.5	48	.64	.72	.07
Matrices	10.8	3.5	22	11.9	3.0	22	.65	.59	.32
Conceptos	11.5	3.0	24	10.4	1.9	24	.25	.25	.45
Figuras incompletas				10.0	2.8	24			
Reconocimiento	9.4	2.7	49						
Localización	10.8	2.6	49						
Búsqueda de animales-Búsqueda de símbolos	10.2	2.4	24	11.0	2.3	24	.33	.40	.37
Cancelación	11.3	3.1	24						
Clave de figuras-Claves	9.4	2.1	23	10.7	2.4	23	.66	.78	.62
Cancelación aleatoria	11.0	3.5	24						
Cancelación estructurada	11.2	2.6	24						
ICV-CIV	102.6	11.5	40	104.5	10.6	40	.54	.64	.17
IVE-CIM	99.7	11.7	44	101.9	14.2	44	.73	.81	.17
IRF-CIM	106.3	17.3	21	106.3	12.8	21	.63	.58	.00
IMT	100.5	11.8	49						
IVP-VP	104.4	12.0	22	105.1	11.5	22	.80	.86	.07
CIT	100.9	12.9	43	109.4	13.1	43	.70	.75	.66
IAV-LG	102.0	13.2	47	109.0	13.5	47	.73	.77	.52
INV	100.9	13.1	40						
ICG	102.9	12.9	36						
ICC	104.6	12.2	24						

Nota: Las abreviaciones en la WPPSI-II son: CIV = CI verbal; CIM = CI manipulativo; VP = Velocidad de procesamiento; LG = Lenguaje general. Las correlaciones se corrigieron de acuerdo con la variabilidad de la muestra de tipificación española de la WPPSI-IV-IV (Guilford y Fruchter, 1978).

[a] Los valores de las columnas «Media» y «*Dt*» son el promedio de las medias y desviaciones típicas de los dos órdenes de aplicación.

[b] La media del coeficiente de correlación de las dos aplicaciones se obtuvo mediante la transformación *z* de Fisher.

[c] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Las medias de los índices primarios de la WPPSI-IV oscilan entre 99.7 (IVE) y 106.3 (IRF), y la media del CI total es de 100.9. Las medias de las puntuaciones compuestas de la WPPSI-III son significativamente superiores a las de las puntuaciones correspondientes de la WPPSI-IV.

Excepto el moderado tamaño del efecto detectado en las diferencias entre las medias del CI total y el IAV-LG, los tamaños del efecto para el resto de los índices fueron insignificantes. Las diferencias entre las medias de las puntuaciones compuestas de la WPPSI-III y la WPPSI-IV siguen la dirección y la magnitud predicha por el efecto Flynn (Flynn, 2007). En cuanto a las

pruebas, se observan un efecto moderado en *Dibujos* y *Clave de figuras-Claves* y efectos pequeños en *Conceptos, Comprensión, Búsqueda de animales-Búsqueda de símbolos* y *Matrices.*

Los coeficientes de correlación corregidos de las puntuaciones compuestas de ambas escalas oscilan entre .58 para el IRF-CIM y .86 para el IVP-VP.

Las correlaciones corregidas entre las pruebas comunes a ambos instrumentos oscilaron entre .25 y .90. *Conceptos* es la prueba que muestra la correlación más baja. Esta prueba se revisó considerablemente para evitar orientar involuntariamente al niño sobre las respuestas correctas; para ello se añadieron estímulos nuevos en algunos ítems. Las pruebas nuevas de Velocidad de procesamiento de la WPPSI-IV se compararon con las pruebas de Velocidad de procesamiento de la WPPSI-III conceptualmente similares (*Búsqueda de animales-Búsqueda de símbolos* y *Clave de figuras-Claves*). Las correlaciones entre estas pruebas son lo suficientemente altas para indicar que las pruebas miden constructos similares.

Estos resultados son consistentes con las investigaciones previas llevadas a cabo para analizar la relación entre las distintas versiones de escalas de inteligencia de Wechsler (Wechsler, 1989, 1991, 1997, 2002, 2003, 2008). La magnitud de las correlaciones y la media consistente de las puntuaciones compuestas sugieren que ambos instrumentos miden constructos similares.

La tabla 6.6 muestra los rangos de las puntuaciones compuestas esperadas para la WPPSI-IV correspondientes a determinadas puntuaciones compuestas de la WPPSI-III (intervalos de confianza del 95 %). Para conocer el rango esperado de una puntuación compuesta de la WPPSI-IV correspondiente a una puntuación compuesta de la WPPSI-III, debe localizarse, en la columna izquierda de la tabla, el valor de la puntuación compuesta de la WPPSI-III deseado y seguir la fila hasta la columna en la que se encuentra el rango de la puntuación compuesta correspondiente de la WPPSI-IV. En muchos casos, todo el rango de puntuaciones esperadas para la WPPSI-IV es inferior a la puntuación de la WPPSI-III. Los rangos de las puntuaciones compuestas esperadas para la WPPSI-IV relacionadas con determinadas puntuaciones compuestas de la WPPSI-III son relativamente pequeños cuando están situados cerca del centro de la distribución de la puntuación (100), mientras que son más amplios en los extremos de los niveles de puntuación.

Tabla 6.6. Rangos de las puntuaciones esperadas para la WPPSI-IV correspondientes a determinadas puntuaciones compuestas de la WPPSI-III (datos españoles)

	Rango de puntuaciones compuestas de la WPPSI-IV					
Puntuación compuesta de la WPPSI-III	**ICV**	**IVE**	**IRF**	**IVP**	**CIT**	**IAV**
55	45-62	48-58	45-73	46-62	40-54	45-55
70	61-75	64-71	57-83	63-75	56-67	58-68
85	79-87	81-85	76-93	81-88	73-80	75-82
100	96-100	96-99	95-105	97-101	89-94	91-95
115	110-116	111-115	109-120	112-117	104-108	106-110
130	123-133	125-131	121-139	125-134	118-125	120-126
145	135-151	138-147	131-155	138-151	131-142	133-143

Nota: El rango del ICV se predijo a partir del CIM de la WPPSI-III, los rangos del IVE y el IRF se predijeron a partir del CIM de la WPPSI-III, el rango del IVP se predijo a partir de la VP de la WPPSI-III, y el rango del IAV se predijo a partir del LG de la WPPSI-III. Los rangos están calculados con un intervalo de confianza del 95% a partir de la media equiparada (Angoff, 1984, Design II.B) de la aplicación de ambas escalas en orden contrabalanceado.

* El rango se ha truncado de acuerdo con las puntuaciones mínimas que pueden obtenerse en los índices.

Correlaciones con las Bayley-III

Las Bayley-III es un instrumento clínico de aplicación individual para evaluar el funcionamiento evolutivo de los niños de 1 a 42 meses. De las escalas que componen las Bayley-III se aplicaron tres: Cognitiva, Lenguaje y Motora. Cada escala ofrece la puntuación compuesta correspondiente. A partir de estudios anteriores (Bayley, 2005; Wechsler, 2002), se esperaba que la puntuaciones compuestas de la WPPSI-IV mostraran correlaciones más altas con la escala Cognitiva que con la escala Motora de las Bayley-III. Asimismo se esperaba que las correlaciones del ICV y el IAV de la WPPSI-IV con la escala Lenguaje fueran más altas que con la escala Cognitiva de las Bayley-III, y que el CI total de la WPPSI-IV mostrara correlaciones iguales o más altas con la escala Lenguaje que con la escala Cognitiva de las Bayley-III. También se predijo que las pruebas *Dibujos* y *Nombres* de la WPPSI-IV mostrarían correlaciones más altas con la escala Lenguaje de las Bayley-III dado que los requisitos de las tareas son similares. Asimismo se esperaba que, de todas las puntuaciones de los índices primarios de la WPPSI-IV, el IVE mostrara la correlación más alta con la escala Motora de las Bayley-III; y que las pruebas *Cubos* y *Localización*, dados los requisitos motores de las tareas, mostraran una correlación más alta con la escala Motora que con otras pruebas de la WPPSI-IV. Por el contrario, considerando resultados anteriores, no se predijo que la escala Motora mostraría una correlación más alta con *Rompecabezas* que con otras pruebas de la WPPSI-IV (Wechsler, 2002).

Se aplicó la WPPSI-IV y las Bayley-III a 115 niños de 2:6 a 3:6 años, en orden contrabalanceado, con un intervalo de aplicación de 9 a 43 días (media = 21 días). La tabla 6.7 muestra las medias, las desviaciones típicas y los coeficientes de correlación corregidos.

Tabla 6.7. Correlaciones entre la WPPSI-IV y las Bayley-III

Prueba/puntuación compuesta de la WPPSI-IV	Puntuaciones compuestas de las Bayley-III			WPPSI-IV		
	Cognitiva	Lenguaje	Motora	Media[a]	*Dt*[a]	*n*
Dibujos	.47	.53	.23	10.0	2.6	114
Información	.63	.65	.28	9.9	2.7	114
Nombres	.53	.64	.14	9.8	2.9	114
Cubos	.41	.40	.24	10.4	2.2	113
Rompecabezas	.48	.46	.10	9.9	3.0	114
Reconocimiento	.50	.44	.16	10.3	2.7	113
Localización	.25	.38	.27	10.6	2.7	114
ICV	.65	.69	.31	100.0	12.3	114
IVE	.58	.56	.21	100.9	12.1	113
IMT	.46	.50	.24	102.5	13.4	113
CIT	.72	.72	.29	101.5	12.0	112
IAV	.56	.66	.20	99.2	13.5	114
INV	.60	.61	.26	102.8	12.5	112
ICG	.71	.71	.30	101.0	11.9	113
Bayley-III						
Media[a]	101.9	104.8	103.9			
Dt[a]	10.1	11.9	14.7			
n	114	114	111			

Nota: Las correlaciones se calcularon por separado para cada orden de aplicación en un diseño contrabalanceado y se corrigieron de acuerdo con la variabilidad de la muestra de tipificación estadounidense de la WPPSI-IV (Guilford y Fruchter, 1978). La media del coeficiente de correlación de las dos aplicaciones se obtuvo mediante la transformación z de Fisher.

[a] Los valores de las columnas «Media» y «*Dt*» son el promedio de las medias y desviaciones típicas de los dos órdenes de aplicación.

Las medias de los índices primarios de la WPPSI-IV oscilan entre 100.0 (ICV) y 102.5 (IMT), y la media del CI total es de 101.5. La media de la escala Cognitiva de las Bayley-III es de 101.9, y las medias de las puntuaciones de las escalas Lenguaje y Motora son de 104.8 y 103.9, respectivamente. Estas medias son ligeramente superiores a las puntuaciones compuestas de la WPPSI-IV, pero las diferencias no son estadísticamente significativas, de acuerdo con el análisis multivariante de la varianza de medidas repetidas ($F[1, 110] = .46$, $p = .50$).

Como se esperaba, las puntuaciones compuestas de la WPPSI-IV muestran correlaciones más altas con las escalas Cognitiva y Lenguaje que con la escala Motora de las Bayley-III. De todas las puntuaciones compuestas de la WPPSI-IV, el CI total muestra la correlación más alta con las escalas Cognitiva y Lenguaje de las Bayley-III (.72). Las correlaciones de las puntuaciones de los índices primarios y secundarios de la WPPSI-IV con la escala Cognitiva de las Bayley-III son moderadamente altas. El IVE muestra una correlación más alta con la escala Cognitiva de las Bayley-III, y el CI total y el ICG muestran la misma correlación tanto con la escala Cognitiva como con la escala Lenguaje. Las puntuaciones de los índices de la WPPSI-IV que están claramente relacionados con el lenguaje (ICV e IAV) muestran correlaciones más altas con la escala Lenguaje que con la puntuación compuesta de la escala Cognitiva de las Bayley-III. Excepto el IAV, las correlaciones de las puntuaciones compuestas de la WPPSI-IV con las escalas Cognitiva y Lenguaje de las Bayley-III son muy similares, lo que refleja la estrecha relación que existe entre el desarrollo del lenguaje y la aptitud cognitiva en este rango de edad (Ford et al., 2012). Cabe destacar que la mayoría de las puntuaciones de la WPPSI-IV muestran correlaciones de bajas a moderadas (.10-.31) con la escala Motora. La correlación entre el INV de la WPPSI-IV y la escala Lenguaje fue ligeramente más alta de lo esperado. No obstante, el INV no es una medida completamente no verbal; más bien se corresponde conceptualmente con una estimación de la aptitud general en la que interviene en menor medida el lenguaje (véase el anexo B de este manual). Todas las pruebas que contribuyen al INV se presentan verbalmente y, dado que el niño debe comprender las consignas de las pruebas, consecuentemente, dichas pruebas requieren la intervención de la aptitud del lenguaje receptivo.

Como se esperaba, las puntuaciones de *Dibujos* y *Nombres* muestran correlaciones más altas con la escala Lenguaje de las Bayley-III que con la escala Cognitiva, al igual que las puntuaciones de *Información*. Asimismo, en relación con la mayoría de las pruebas de la WPPSI-IV, *Cubos* y *Localización* muestran correlaciones más altas con la escala Motora de las Bayley-III (.24 y .27, respectivamente). La correlación no demasiado alta (.10) entre *Rompecabezas* y la escala Motora es la que se esperaba, De todas las puntuaciones de las pruebas de la WPPSI-IV, las correlaciones más altas con la escala Cognitiva de las Bayley-III se observan en *Información, Nombres, Reconocimiento, Rompecabezas* y *Dibujos*. Por lo general, los resultados son consistentes con las investigaciones previas (Bayley, 2005; Wechsler, 2002). La consistencia de los resultados y la magnitud de las correlaciones ratifican la validez de constructo de la WPPSI-IV para los rangos de menor edad.

La tabla 6.8 muestra los rangos de las puntuaciones compuestas esperadas para la WPPSI-IV correspondientes a determinadas puntuaciones compuestas de las Bayley-III (intervalos de confianza del 95 %). Para conocer el rango esperado de una puntuación compuesta de la WPPSI-IV correspondiente a una puntuación compuesta de las Bayley-III, debe localizarse, en la columna izquierda de la tabla, el valor de la puntuación compuesta de las Bayley-III deseado y seguir la fila hasta la columna en la que se encuentra el rango de la puntuación compuesta correspondiente de la WPPSI-IV. En algunos casos, todo el rango de puntuaciones esperadas para la WPPSI-IV es inferior a la puntuación compuesta de las Bayley-III. Al igual que en el caso de la WPPSI-III, y por motivos similares, los rangos de las puntuaciones compuestas esperadas para la WPPSI-IV relacionadas con determinadas puntuaciones compuestas de las Bayley-III son relativamente pequeños cuando están situados cerca del centro de la distribución de la puntuación (100), mientras que son más amplios en los extremos de los niveles de puntuación.

Tabla 6.8. Rangos de las puntuaciones esperadas para la WPPSI-IV correspondientes a determinadas puntuaciones compuestas de las Bayley-III

	Rango de puntuaciones compuestas de la WPPSI-IV				
Puntuaciones compuestas de la Bayley-III	***ICV* según punt. escala Cognitiva Bayley-III**	***IVE* según punt. escala Cognitiva Bayley-III**	***CIT* según punt. escala Cognitiva Bayley-III**	***IAV* según punt. escala Lenguaje Bayley-III**	**ICG según punt. escala Cognitiva Bayley-III**
55	45–61	45–63	47–62	45[a]–57	47–62
70	62–74	63–75	64–75	59–70	64–74
85	80–87	80–88	81–88	76–83	81–87
100	96–100	97–101	98–101	92–97	97–101
115	110–116	111–117	112–117	107–112	112–117
130	123–133	124–134	125–134	120–129	125–134
145	136–151	136–152	138–151	133–146	137–151

Nota: Los rangos están calculados con un intervalo de confianza del 95 % a partir de la media equiparada (Angoff, 1984, Design II.B) de los datos de la aplicación de ambas escalas en orden contrabalanceado.

[a] El rango se ha truncado de acuerdo con las puntuaciones mínimas que pueden obtenerse en los índices.

Correlaciones con la WISC-IV

La WISC-IV es un instrumento clínico de aplicación individual diseñado para medir la aptitud cognitiva de los niños de 6 años a 16 años 11 meses. La WISC-IV ofrece puntuaciones de las pruebas y puntuaciones de los índices que reflejan el funcionamiento intelectual en determinadas áreas cognitivas: comprensión verbal (CV), razonamiento perceptivo (RP), memoria de trabajo (MT) y velocidad de procesamiento (VP), así como una puntuación compuesta que representa la aptitud intelectual general (CI total). En estudios posteriores (Raiford et al., 2005; Weiss y Gabel, 2008; Weiss, Saklofske, Prifitera y Holdnack, 2006) se proporcionó un índice de capacidad general (ICG) y un índice de competencia cognitiva (ICC).

Se aplicó la WPPSI-IV y la WISC-IV a 99 niños de 6:6 a 7:7 años, en orden contrabalanceado, con un intervalo de aplicación de 14 a 54 días (media = 22 días). La tabla 6.9 muestra las medias, las desviaciones típicas, las correlaciones (corregidas y sin corregir) y las diferencias típicas.

Tabla 6.9. Correlaciones entre la WPPSI-IV y la WISC-IV

Prueba/puntuación de procesamiento/ puntuación compuesta	WPPSI-IV			WISC-IV					
	Media[a]	*Dt*[a]	*n*	Media[a]	*Dt*[a]	*n*	r_{12}[b]	r_{12}[b] corregido	Diferencia típica[c]
Información	9.8	2.3	95	10.5	2.5	95	.65	.71	.29
Semejanzas	9.8	2.3	97	11.1	2.5	97	.67	.74	.54
Vocabulario	9.7	2.6	95	10.0	2.2	95	.71	.76	.12
Comprensión	9.2	2.5	97	9.6	2.6	97	.53	.59	.16
Dibujos	10.4	2.7	99						
Nombres	9.9	2.6	99						
Adivinanzas				10.4	2.4	99			
Cubos	10.3	2.9	97	11.0	2.7	97	.61	.64	.25
Rompecabezas	10.1	2.4	99						
Matrices	10.3	2.7	97	11.4	2.8	97	.62	.67	.40
Conceptos	10.5	2.1	92	11.7	3.1	92	.53	.61	.45
Figuras incompletas				9.6	3.0	99			
Reconocimiento	9.8	2.9	99						
Localización	10.6	2.5	99						
Dígitos				10.0	2.7	99			
Letras y números				10.6	2.6	99			
Aritmética				9.9	2.4	99			
Búsqueda de animales-Búsqueda de símbolos	10.0	2.7	93	11.4	2.6	93	.48	.53	.53
Cancelación	9.8	2.5	92	10.2	2.5	92	.50	.58	.16
Clave de figuras-Claves	10.1	2.9	98	10.5	3.0	98	.56	.56	.14
Cancelación aleatoria	10.0	2.6	95	10.0	2.5	95	.51	.59	.00
Cancelación estructurada	9.8	2.4	92	10.4	2.7	92	.38	.42	.24
ICV-CV	98.7	11.6	93	101.0	10.6	93	.63	.73	.21
IVE-RP	101.8	12.3	96	108.6	13.6	96	.67	.75	.52
IRF-RP	102.9	10.3	89	109.1	13.7	89	.70	.77	.51
IMT-MT	101.1	13.4	99	101.1	12.7	99	.50	.55	.00
IVP-VP	99.6	12.7	86	102.9	13.6	86	.56	.60	.25
CIT	100.2	12.7	87	105.3	12.5	87	.80	.84	.40
IAV	100.4	12.2	99						
INV	101.7	11.6	86						
ICG	100.3	12.3	90	105.7	11.8	90	.78	.84	.45
ICC	100.1	12.3	86	104.0	13.0	86	.70	.74	.31

Nota: Las abreviaciones en la WISC-IV son: CV = Comprensión verbal; RP = Razonamiento perceptivo; MT = Memoria de trabajo; VP = Velocidad de procesamiento. Las correlaciones se calcularon por separado para cada orden de aplicación en un diseño contrabalanceado) y se corrigieron de acuerdo con la variabilidad de la muestra de tipificación estadounidense de la WPPSI-IV (Guilford y Fruchter, 1978).

[a] Los valores de las columnas «Media» y «*Dt*» son el promedio de las medias y desviaciones típicas de los dos órdenes de aplicación.

[b] La media del coeficiente de correlación de las dos aplicaciones se obtuvo mediante la transformación z de Fisher.

[c] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Las medias de los índices primarios de la WPPSI-IV oscilan entre 98.7 (ICV) y 102.9 (IRF), y la media del CI total es de 100.2. Con excepción de MT, las puntuaciones compuestas de la WISC-IV (con una puntuación comparable en la WPPSI-IV) son consistentemente más altas que las puntuaciones de los índices primarios de la WPPSI-IV. Sin embargo, estas diferencias no son estadísticamente significativas, de acuerdo con el análisis multivariante de la varianza de medidas repetidas (F [6, 414] = 1.56, p = .16). Estos resultados son consistentes con los obtenidos en estudios previos (Wechsler, 1991, 2002). Con excepción de MT, todas las diferencias muestran tamaños del efecto de pequeños a moderados, siendo los mayores los observados en las comparaciones IVE-RP e IRF-RP. Esto no debe sorprender si se tiene en cuenta que las pruebas que conforman el IVE y el IRF de la WPPSI-IV difieren de las que componen el RP de la WISC-IV.

Los tamaños del efecto observados en las pruebas son moderados en *Semejanzas* y *Búsqueda de animales-Búsqueda de símbolos* y pequeños en *Información, Cubos, Matrices* y *Conceptos.*

Los coeficientes de correlación corregidos entre los índices primarios de la WPPSI-IV y los índices de la WISC-IV oscilan entre .55 (IMT-MT) y .77 (IRF-RP), y el de la correlación CI total es de .84. La correlación entre el ICG y el ICC de ambas escalas es alta (.84 y .74, respectivamente). La correlación entre el IMT de la WPPSI-IV y el MT de la WISC-IV demuestra la validez de esta puntuación compuesta nueva. Dado que el IMT de la WPPSI-IV incluye tareas visuales y visoespaciales, mientras que el MT de la WISC-IV solo incluye tareas verbales, la correlación entre estos índices es superior a la esperada y sugiere que ambos miden un constructo similar.

Respecto a las pruebas, los coeficientes de correlación corregidos oscilan entre .53 (*Búsqueda de animales-Búsqueda de símbolos*) y .76 (*Vocabulario*). En general, las correlaciones de las pruebas de Comprensión verbal de la WPPSI-IV con las pruebas de Comprensión verbal de la WISC-IV son las más altas, les siguen las correlaciones de las pruebas de Visoespacial y las de Razonamiento fluido de la WPPSI-IV con las pruebas de Razonamiento perceptivo de la WISC-IV, y, finalmente, las correlaciones entre las pruebas de Velocidad de procesamiento de ambas escalas. No debe sorprender que la correlación más baja se dé entre las pruebas de Velocidad de procesamiento puesto que las pruebas de ambas escalas son muy diferentes. En conjunto, estos resultados son muy consistentes con los observados entre la WPPSI-III y la WISC-IV (Wechsler, 2003), así como en otros estudios comparativos de distintas ediciones de la WPPSI y la WISC (Wechsler, 1991, 2002). La magnitud de las correlaciones y la media consistente de las puntuaciones compuestas sugieren que ambos instrumentos miden constructos similares.

La tabla 6.10 muestra los rangos de las puntuaciones compuestas esperadas para la WISC-IV correspondientes a determinadas puntuaciones compuestas de la WPPSI-IV (intervalos de confianza del 95 %). Para conocer el rango esperado de una puntuación compuesta de la WISC-IV correspondiente a una puntuación compuesta de la WPPSI-IV, debe localizarse, en la columna izquierda de la tabla, el valor de la puntuación compuesta de la WPPSI-IV deseado y seguir la fila hasta la columna en la que se encuentra el rango de la puntuación compuesta correspondiente de la WISC-IV. Cabe señalar que los rangos de las puntuaciones compuestas esperadas para la WISC-IV relacionadas con determinadas puntuaciones compuestas de la WPPSI-IV son relativamente pequeños cuando están situados cerca del centro de la distribución de la puntuación (100), mientras que son más amplios en los extremos de los niveles de puntuación. Asimismo adviértase que el punto medio del rango de puntuaciones esperadas para la WISC-IV es, en la mayoría de los casos, superior a la puntuación correspondiente de la WPPSI-IV, y que algunos rangos de la WISC-IV no contienen el valor correspondiente de la WPPSI-IV.

Tabla 6.10. Rangos de las puntuaciones compuestas esperadas para la WISC-IV correspondientes a determinadas puntuaciones compuestas de la WPPSI-IV

	Rango de puntuaciones compuestas de la WISC-IV							
Puntuación compuesta de la WPPSI-IV	**ICV**	**RP según IVE WPPSI-IV**	**RP según IRF WPPSI-IV**	**MT**	**VP**	**CIT**	**ICG**	**ICC**
55	52–63	54–69	52–70	50[a]–62	50–67	55–65	55–66	52–66
70	69–76	72–82	70–82	65–75	68–79	72–79	72–79	69–79
85	85–89	89–95	87–94	82–88	85–92	88–92	88–92	86–92
100	101–104	105–109	104–108	98–102	101–106	104–107	104–107	102–106
115	115–120	119–125	118–124	112–118	115–122	118–122	118–123	116–122
130	128–136	132–142	130–142	125–135	127–139	131–139	132–139	129–139
145	141–153	144–155[a]	142–155[a]	138–150[a]	140–150[a]	145–156	145–156	142–156

Nota: Los rangos están calculados con un intervalo de confianza del 95 % a partir de la media equiparada (Angoff, 1984, Design II.B) de los datos de la aplicación de ambas escalas en orden contrabalanceado.

[a] El rango se ha truncado de acuerdo con las puntuaciones que pueden obtenerse en los índices.

Correlaciones con el NNAT2

El NNAT2 es un test de selección múltiple que mide el razonamiento a partir de estímulos figurativos (matrices) en niños de 4 años a 18 años 11 meses. El nivel A, utilizado en este estudio, contiene 48 ítems. El NNAT2 ofrece una única puntuación típica: el Índice de aptitud Naglieri (IAN), con una media de 100 y una desviación típica de 16. Se esperaba que el CI total de la WPPSI-IV y el IAN del NNAT2 mostraran correlaciones relativamente altas, dado que ambas puntuaciones miden la aptitud intelectual general. Asimismo se esperaba que la correlación entre el IAN y la prueba Matrices de la WPPSI-IV fuera moderadamente alta, de acuerdo con estudios previos (Naglieri, 2007) y porque el NNAT2 se diseñó para medir un constructo similar con el mismo formato de ítems.

Se aplicó la WPPSI-IV y el NNAT-2, en este orden, a 62 niños de 4:0 a 4:11 años, con un intervalo de aplicación de 0 a 93 días (media = 26 días). Aunque el NNAT2 se aplica típicamente en grupo, puede aplicarse individualmente a los niños más pequeños. En este estudio se aplicó individualmente. La tabla 6.11 muestra las medias, las desviaciones típicas y los coeficientes de correlación corregidos.

Tabla 6.11. Correlaciones entre la WPPSI-IV y el NNAT2

Prueba/punt. de procesamiento/ punt. compuesta de la WPPSI-IV	NNAT2	WPPSI-IV		
	IAN	Media	*Dt*	*n*
Información	.42	10.2	2.8	62
Semejanzas	.48	10.0	3.0	62
Vocabulario	.41	10.0	2.8	62
Comprensión	.36	9.8	3.2	62
Dibujos	.42	9.7	3.2	62
Nombres	.52	10.2	2.6	62
Cubos	.73	9.7	2.3	62
Rompecabezas	.47	9.4	3.5	62
Matrices	.57	10.2	2.7	62
Conceptos	.31	10.1	3.1	62
Reconocimiento	.54	10.2	2.4	62
Localización	.52	10.2	2.7	62
Búsqueda de animales	.30	10.7	2.8	62
Cancelación	.10	10.6	2.8	62
Clave de figuras	.27	9.9	3.1	62
Cancelación aleatoria	.14	10.5	2.7	62
Cancelación estructurada	.04	10.5	3.0	62
ICV	.48	100.6	14.6	62
IVE	.60	97.9	15.6	62
IRF	.52	100.9	14.2	62
IMT	.65	101.2	12.4	62
IVP	.25	104.0	13.5	62
CIT	.64	101.2	14.1	62
IAV	.53	99.5	14.4	62
INV	.63	101.2	13.8	62
ICG	.64	100.1	14.1	62
ICC	.51	103.1	13.1	62
NNAT2				
Media	104.1			
Dt	13.3			
n	62			

Nota: IAN = Índice de aptitud Naglieri Se corrigieron las correlaciones de acuerdo con la variabilidad de la muestra de tipificación estadounidense de la WPPSI-IV (Guilford & Fruchter, 1978).

Las medias de los índices primarios de la WPPSI-IV oscilan entre 97.9 (IVE) y 104.0 (IVP), y la media del CI total es de 101.2. La media del IAN es de 104.1, lo que probablemente refleja los efectos de aplicación repetida, dado que la NNAT2 siempre se aplicó después de la WPPSI-IV. Como se esperaba, la correlación entre el IAN y el CI total es alta (.64). En cuanto a los índices primarios de la WPPSI-IV, el IMT, el IVE y el IRF muestran correlaciones más altas con el IAN. La alta correlación del IMT con el IAN probablemente refleja la intervención de la memoria de trabajo en las tareas de razonamiento, así como la implicación del córtex prefrontal en las medidas de memoria de trabajo y de razonamiento. No debe sorprender la estrecha relación mostrada por el IVE y el IAN puesto que las pruebas que contribuyen al IVE se basan, al igual que el NNAT2, en estímulos abstractos. La fuerte correlación del IRF con el IAN refleja la subyacente aptitud de razonamiento fluido inherente a ambas pruebas. En cuanto a los índices secundarios, el INV y el ICG muestran correlaciones moderadamente altas con el IAN.

Respecto a las pruebas, las correlaciones más altas con el IAN se observan en *Cubos, Matrices, Reconocimiento, Nombres* y *Localización*. Estos resultados son los esperados, puesto que tanto *Cubos* y *Localización* como el IAN se basan en estímulos presentados espacialmente. Por su parte, la alta correlación entre *Matrices* y el IAN se debe a que en ambos interviene el razonamiento fluido y a que el formato de presentación de los estímulos (matrices) es similar. En cuanto a las altas correlaciones observadas en *Reconocimiento* y *Nombres* con el IAN, pueden deberse a la relación que comparten con la inteligencia general. Además, en particular, la correlación entre *Reconocimiento* y el IAN puede deberse a la importancia que tiene la memoria de trabajo para el razonamiento fluido. Los resultados de este estudio demuestran que la WPPSI-IV y el NNAT2 miden constructos relacionados con la aptitud intelectual.

Correlaciones con la NEPSY-II

La NEPSY-II es un instrumento exhaustivo diseñado para medir el desarrollo neuropsicológico en niños en edad preescolar y escolar. Los resultados obtenidos de la evaluación con la NEPSY-II aportan datos para el diagnóstico de diversos trastornos de la infancia y son de ayuda para planificar la intervención correspondiente. La NEPSY-II ofrece puntuaciones de las pruebas, no puntuaciones compuestas. Esta escala esta formada por una amplia gama de pruebas que se reparten entre seis dominios funcionales. En este estudio se utilizan determinadas pruebas pertenecientes a cuatro de los seis dominios funcionales de la NEPSY-II, seleccionadas por su relevancia en los motivos de consulta y en los trastornos clínicos de la primera infancia. Las pruebas seleccionadas fueron: *Reconocimiento de emociones* y *Teoría de la mente*, del dominio Percepción social; *Memoria de caras, Memoria de diseños* y *Memoria narrativa*, del dominio Memoria y aprendizaje; *Procesamiento fonológico* y *Velocidad de denominación*, del dominio Lenguaje; e *Inhibición, Atención auditiva* y *Estatua*, del dominio Atención y función ejecutiva. *Memoria de caras* mide un tipo de memoria muy específico y también se considera una prueba del dominio Percepción social. Las correlaciones entre la WPPSI-IV y la NEPSY-II aportan datos sobre la validez convergente y discriminante de las pruebas de la WPPSI-IV.

Considerando estudios previos sobre las relaciones de la NEPSY-II con la WISC-IV (Korkman et al., 2007), se predijo que las puntuaciones de las pruebas y las puntuaciones compuestas de la WPPSI-IV mostrarían correlaciones más altas con las pruebas de la NEPSY-II cuyos procesos cognitivos subyacentes, formas de presentación de los estímulos y tipos de respuestas requeridas fueran similares (p. ej., visuales frente a verbales).

Se aplicó la WPPSI-IV y las pruebas seleccionadas de la NEPSY-II a 55 niños de 3:0 a 7:6 años, con un intervalo de aplicación de 0 a 45 días (media = 11 días). En todos los casos se aplicó la WPPSI-IV antes que la NEPSY-II. La tabla 6.12 muestra las medias, las desviaciones típicas y los coeficientes de correlación corregidos.

Tabla 6.12. Correlaciones entre la WPPSI-IV y la NEPSY-II

	Pruebas de la NEPSY-II									
Prueba/punt. de procesamiento/ punt. compuesta de la WPPSI-IV	**RE**	**TM**	**MC**	**MD Contenido**	**MD Espacial**	**MD**	**MN Rec. libre**	**PF**	**Total de tiempo empleado VD**	**Combinada VD**
Información	.21	.51	.03	−.02	.06	−.04	.60	.34	.32	.46
Semejanzas	.21	.38	.20	−.07	−.04	−.27	.56	.48	−.10	.13
Vocabulario	.33	.50	.26	−.01	.20	.00	.70	.53	.26	.40
Comprensión	.28	.56	.07	.25	.22	.26	.41	.62	.45	.52
Dibujos	.20	.30	.08	−.18	−.07	−.11	.34	.49	.35	.42
Nombres	.22	.25	.02	−.13	−.21	−.21	.45	.49	.30	.36
Cubos	.30	.20	.23	.08	.17	.27	.42	.42	.41	.45
Rompecabezas	.43	.13	.47	.08	.14	.21	.45	.32	.32	.34
Matrices	.15	.21	.44	.15	.23	.18	.33	.37	.48	.48
Conceptos	.47	.15	.51	.04	.15	.09	.44	.36	.31	.35
Reconocimiento	.33	.42	.22	.21	.18	.26	.48	.51	.31	.42
Localización	.27	.29	.28	.09	.17	.18	.37	.42	.16	.31
Búsqueda de animales	.19	.32	.18	.05	.20	.13	.36	.35	.38	.42
Cancelación	.13	.02	.25	.12	.30	.20	.50	.03	.35	.42
Clave de figuras	.33	.20	.39	−.10	.09	−.08	.56	.21	.35	.35
Cancelación aleatoria	.23	.10	.27	.13	.24	.22	.45	.17	.37	.43
Cancelación estructurada	−.01	−.08	.11	.10	.24	.11	.51	−.10	.26	.30
ICV	.24	.46	.12	−.10	.05	−.12	.65	.41	.22	.43
IVE	.48	.21	.44	.09	.19	.30	.54	.48	.47	.51
IRF	.33	.20	.52	.10	.20	.14	.41	.41	.43	.45
IMT	.34	.39	.30	.16	.19	.24	.49	.52	.25	.41
IVP	.19	.23	.25	.10	.30	.19	.51	.26	.44	.49
CIT	.37	.47	.34	.11	.25	.23	.67	.59	.52	.65
IAV	.26	.35	.06	−.18	−.16	−.18	.47	.59	.39	.48
INV	.38	.37	.43	.14	.25	.27	.54	.53	.53	.60
ICG	.38	.48	.38	.07	.25	.20	.67	.56	.54	.66
ICC	.36	.37	.33	.19	.29	.27	.60	.50	.36	.49
NEPSY-II										
Media	11.1	9.7	11.1	11.3	11.0	11.4	10.0	10.7	9.7	10.0
Dt	3.3	2.8	2.9	3.7	3.4	3.8	2.6	3.3	3.4	3.5
n	54	46	29	49	49	49	29	55	54	54

Nota: Las abreviaciones de la NEPSY-II son: RE = *Reconocimiento de emociones*, TM = *Teoría de la mente*, MC = *Memoria de caras*, MD = *Memoria de diseños*, MN = *Memoria narrativa*, PF = *Procesamiento fonológico*, VD = *Velocidad de denominación*. Se corrigieron las correlaciones de acuerdo con la variabilidad de la muestra de tipificación de la WPPSI-IV (Guilford & Fruchter, 1978).

Tabla 6.12. Correlaciones entre la WPPSI-IV y la NEPSY-II *(continuación)*

	Pruebas de la NEPSY-II								WPPSI-IV		
Prueba/punt. de procesamiento/ punt. compuesta de la WPPSI-IV	**Total de tiempo empleado IND**	**Combinada IND**	**Total de tiempo empleado INI**	**Combinada INI**	**Total errores IN**	**Total resp. correctas AA**	**Combinada AA**	**Estatua**	**Media**	***Dt***	***n***
Información	−.14	.13	−.04	.28	.30	.04	.00	.26	10.0	2.8	55
Semejanzas	−.07	.09	.00	.43	.43	.11	.21	.06	10.1	2.1	42
Vocabulario	−.11	−.04	−.02	.30	.34	.19	.15	.32	10.5	2.4	42
Comprensión	.27	.28	.20	.35	.33	.11	.09	.54	9.8	2.4	42
Dibujos	.03	.06	−.09	.28	.30	.07	.05	.27	10.7	2.9	55
Nombres	.13	.01	−.02	.24	.26	.24	.12	.36	9.9	2.6	55
Cubos	.23	.15	.37	.53	.51	.43	.47	.48	11.2	3.0	55
Rompecabezas	−.27	.01	−.20	.10	.29	.21	.18	.22	10.3	3.2	55
Matrices	.02	.29	.04	.24	.26	.32	.39	.21	11.3	3.1	42
Conceptos	−.12	.19	−.10	.24	.37	.45	.48	.35	10.7	3.0	42
Reconocimiento	.19	.20	.23	.43	.34	.50	.54	.28	11.1	3.0	55
Localización	−.07	.03	.15	.33	.34	.20	.28	.06	11.2	3.5	55
Búsqueda de animales	.06	.16	.19	.29	.33	.17	.15	.48	10.6	3.2	42
Cancelación	.07	−.09	.08	.11	.22	−.25	−.23	.22	9.9	2.7	42
Clave de figuras	.07	.16	−.06	−.01	.14	.28	.22	.23	10.2	3.1	42
Cancelación aleatoria	.00	−.12	−.02	.10	.22	−.15	−.13	.31	10.1	3.0	42
Cancelación estructurada	.18	−.06	.21	.10	.14	−.31	−.31	.05	9.7	2.5	42
ICV	−.13	.13	−.05	.38	.40	.11	.12	.17	100.5	12.6	55
IVE	−.03	.09	.11	.40	.50	.41	.41	.44	104.7	14.7	55
IRF	−.06	.26	−.04	.27	.36	.42	.48	.32	106.0	15.7	42
IMT	.04	.10	.21	.43	.39	.38	.46	.17	107.2	16.4	55
IVP	.07	.07	.18	.28	.36	.00	.00	.45	101.7	14.7	42
CIT	.12	.28	.24	.57	.55	.43	.47	.44	104.6	14.1	55
IAV	.08	.06	−.07	.31	.33	.18	.09	.39	101.4	13.2	55
INV	.09	.24	.19	.47	.49	.49	.53	.45	106.0	15.6	55
ICG	.06	.28	.19	.56	.56	.40	.45	.41	103.2	12.9	55
ICC	.06	.11	.24	.43	.45	.25	.29	.34	105.0	15.3	42
NEPSY-II											
Media	10.3	11.4	10.4	10.1	10.3	10.1	10.3	9.4			
Dt	3.1	3.3	3.5	4.1	4.9	3.1	3.6	3.2			
n	28	28	28	28	28	29	29	46			

Nota: Las abreviaciones de la NEPSY-II son: IND = *Inhibición*-Denominación, INI = *Inhibición*-Inhibición, IN = *Inhibición*, AA = *Atención auditiva*. Se corrigieron las correlaciones de acuerdo con la variabilidad de la muestra de tipificación de la WPPSI-IV (Guilford & Fruchter, 1978).

Las medias de los índices de la WPPSI-IV oscilan entre 100.5 (ICV) y 107.2 (IMT), y la media del CI total es de 104.6. Las medias del IMT y del CI total son ligeramente más altas que las de la muestra de tipificación probablemente porque la proporción de niños cuyos padres tenían un nivel educativo alto era superior en la muestra de este estudio. Las correlaciones entre las pruebas de la NEPSY y las puntuaciones de las pruebas y las puntuaciones compuestas de la WPPSI-IV fueron las esperadas, considerando los procesos cognitivos subyacentes, así como el tipo de presentación de los estímulos y de las respuestas. El CI total muestra correlaciones más altas con *Memoria narrativa*-Recuerdo libre, con la puntuación escalar combinada de *Velocidad de denominación* y con *Procesamiento fonológico* de la NEPSY-II, lo que refleja la importancia que tiene el desarrollo del lenguaje para la aptitud intelectual en este rango de edad. Además, el CI total muestra correlaciones moderadas con la puntuación escalar combinada de *Inhibición*-Inhibición y el total de errores de *Inhibición*. Estos resultados son los esperados, dado que el control cognitivo es fundamental para la expresión de las aptitudes intelectuales en este rango de edad, y la aptitud cognitiva general es un sólido indicador del desarrollo del control inhibitorio a lo largo de la edad preescolar (Wiebe, Sheffield y Espy, 2012).

El ICV muestra la correlación más alta con *Memoria narrativa*-Recuerdo libre de la NEPSY-II, lo que refleja que comparten el formato verbal de la presentación y de las respuestas. Como se esperaba, las correlaciones del ICV con *Procesamiento fonológico* y la puntuación escalar combinada de *Velocidad de denominación* son moderadas. Las correlaciones moderadas entre el ICV y algunas puntuaciones de las pruebas *Inhibición* y *Teoría de la mente* fueran las esperadas, puesto que los estudios demuestran los vínculos existentes entre las aptitudes verbales, el desarrollo de la teoría de la mente y las funciones ejecutivas (p. ej., la inhibición) en los niños pequeños (Bjorklund, Cormier y Rosenberg, 2005; Schneider, Lockl y Fernandez, 2005; Tager-Flusberg y Joseph, 2005). Las correlaciones del ICV con las tareas de Memoria y aprendizaje en las que intervienen estímulos visuales y espaciales y con las puntuaciones de *Inhibición* que miden el tiempo empleado son más bajas. Las pruebas de Comprensión verbal de la WPPSI-IV muestran correlaciones de moderadas a altas con *Memoria narrativa, Teoría de la mente* y las pruebas del dominio Lenguaje de la NEPSY-II; y correlaciones más bajas con las puntuaciones de *Inhibición* que miden el tiempo empleado.

El IVE muestra una correlación más alta con *Memoria narrativa*-Recuerdo libre que con el resto de las puntuaciones de la NEPSY-II. Las correlaciones del IVE con las puntuaciones de *Velocidad de denominación* y con el total de errores de *Inhibición* son moderadas. Asimismo, la correlación entre el IVE y *Procesamiento fonológico* es moderadamente alta. La estrecha relación observada entre el IVE y las pruebas *Velocidad de denominación* y *Procesamiento fonológico* no debe sorprender puesto que recientemente se han vinculado las destrezas previas a la lectoescritura y las destrezas lectoras, como las que miden las pruebas de la NEPSY-II, con el desarrollo de la aptitud visoespacial en los niños pequeños (McBride-Chang et al., 2011). Otra demostración de la relación entre el IVE y las destrezas previas a la lectoescritura la aportó un estudio realizado con niños con problemas previos a la lectoescritura y con un grupo control; este estudio, descrito más adelante en este capítulo, señaló que, al comparar las medias de los índices primarios de la WPPSI-IV, el IVE mostraba el mayor tamaño del efecto para las diferencias entre las medias de los grupos. El IVE muestra una correlación más baja con el total de tiempo empleado de *Inhibición*-Denominación, lo que sugiere que las funciones cognitivas de nivel superior están más relacionadas con los aspectos reguladores del rendimiento en la prueba *Inhibición* que con el componente de identificación rápida de la tarea. La prueba *Cubos* muestra correlaciones más altas con las siguientes puntuaciones de las pruebas del dominio Inhibición de la NEPSY-II: *Inhibición, Estatua* y *Atención auditiva*. En estas pruebas interviene la capacidad de inhibir errores y la de mantener la atención durante cierto tiempo, capacidades que pueden dar lugar a un mejor rendimiento en *Cubos*. Las aptitudes espaciales se habían relacionado anteriormente con el control inhibitorio (Atkinson et al., 2003). La prueba *Rompecabezas* está más relacionada con *Memoria de caras, Memoria narrativa*-Recuerdo libre y *Reconocimiento de emociones*. Aún así, la relación de *Rompecabezas* con *Reconocimiento de emociones* y *Memoria de caras* era relativamente esperada, porque en estas tres pruebas interviene el procesamiento visual de estímulos significativos, y los estudios basados en imágenes cerebrales vinculan determinados aspectos de la percepción facial con la destreza visual (Van Bavel, Packer y Cunningham, 2011). La relación entre *Rompecabezas* y *Memoria narrativa* probablemente refleja que ambas están asociadas a la aptitud intelectual general.

El IRF muestra correlaciones moderadamente altas con *Memoria de caras, Memoria narrativa*-Recuerdo libre, *Procesamiento fonológico, Velocidad de denominación* y *Atención auditiva* de la NEPSY-II. En estudios anteriores realizados con la NEPSY-II también se observaron relaciones similares con puntuaciones de pruebas y puntuaciones compuestas de razonamiento fluido (Korkman et al., 2007). La relación del IRF con *Memoria de caras* probablemente refleja que en ambos interviene el procesamiento visual de estímulos significativos. La correlación moderada entre el IRF y *Memoria narrativa* puede indicar que ambos guardan mucha relación con *g*. Las asociaciones moderadas observadas entre el IRF y las pruebas *Procesamiento fonológico* y *Velocidad de denominación*, del dominio Lenguaje de la NEPSY-II, no se esperaban. El razonamiento fluido se ha asociado a los problemas de comprensión lectora, pero no consistentemente a los problemas de procesamiento fonológico o de denominación automatizada rápida, relacionados con otros tipos de trastornos de la lectura. Sin embargo, las destrezas lingüísticas pueden mejorar el rendimiento en tareas de razonamiento visual, en especial en las que requieren conocimiento semántico, como la prueba *Conceptos* de la WPPSI-IV. Además, en estudios anteriores (Korkman et al., 2007) se habían observado correlaciones similares entre tareas de razonamiento fluido y estas pruebas de la NEPSY-II; por ello, es necesario investigar esta asociación con más detalle. La relación entre el IRF y la prueba *Atención auditiva*, del dominio Atención y función ejecutiva, se esperaba en parte, dado que tanto el razonamiento fluido como la atención se conceptualizan como procesos ejecutivos. Las correlaciones más altas entre las pruebas de Razonamiento fluido y las pruebas de la NEPSY-II son similares a las observadas entre el IRF y las pruebas de la NEPSY-II. La prueba *Matrices* muestra las correlaciones más altas con *Memoria de caras* y *Velocidad de denominación*. Las correlaciones más altas de la prueba *Conceptos* se observan con *Memoria de caras, Atención auditiva, Reconocimiento de emociones* y *Memoria narrativa*. Estas pruebas muestran correlaciones más bajas con las puntuaciones de tiempo empleado de la prueba *Inhibición*.

El IMT muestra correlaciones más altas con *Procesamiento fonológico, Memoria narrativa*-Recuerdo libre, puntuación escalar combinada de *Atención auditiva, Velocidad de denominación* y las puntuaciones escalares combinadas de *Inhibición* (IND e INI) de la NEPSY-II. Estos resultados son esperados en parte, puesto que el funcionamiento de la memoria es importante para la memoria inmediata. La relación entre el IMT y la prueba *Memoria narrativa* probablemente se debe a que ambos guardan una estrecha relación con la aptitud intelectual general. Respecto a la relación del IMT con la prueba *Velocidad de denominación*, se ha asociado la memoria de trabajo con el desarrollo del lenguaje (Archibald y Gathercole, 2006a) y con los trastornos de la lectura, asociados a problemas de denominación automatizada rápida (Flanagan, Alfonso, Mascolo y Sotelo-Dynega, 2012). La relación entre el IMT y la prueba *Atención auditiva* no debe sorprender, puesto que algunos modelos de la memoria de trabajo consideran que esta incluye el enfoque atencional (Cowan, 1988; Towse y Cowan, 2005). Además, tanto la memoria de trabajo como el control inhibitorio se clasifican como funciones ejecutivas, y distintos estudios de análisis factorial confirman que la memoria de trabajo y el control inhibitorio muestran una estrecha relación en los niños pequeños (Bull, Espy, Wiebe, Sheffield y Nelson, 2011; Wiebe, Epsy y Charak, 2008). Las correlaciones más altas entre las pruebas de Memoria de trabajo y las pruebas de la NEPSY-II son similares a las observadas entre el IMT y las pruebas de la NEPSY-II. Las pruebas *Reconocimiento* y *Localización* muestran correlaciones moderadamente altas con *Memoria narrativa*-Recuerdo libre y *Procesamiento fonológico*. El rendimiento en ambas pruebas de la NEPSY-II mejora cuando la memoria de trabajo es buena. *Reconocimiento* también muestra una correlación moderada con *Atención auditiva*.

El IVP muestra correlaciones más altas con las pruebas *Memoria narrativa* y *Velocidad de denominación* de la NEPSY-II. Estos resultados son muy similares a los observados entre VP de la WISC-IV y la NEPSY-II (Korkman et al., 2007). Las correlaciones más altas entre las pruebas de Velocidad de procesamiento y las pruebas de la NEPSY-II son similares a las correlaciones más altas observadas entre el IVP y las pruebas de la NEPSY-II. La correlación más alta de la prueba *Búsqueda de animales* es moderadamente alta y se observa con la prueba *Estatua*; este resultado puede estar relacionado con el hecho de que ambas requieren inhibir una respuesta motora. Las pruebas *Cancelación* y *Clave de figuras* muestran las correlaciones más altas con *Memoria narrativa*, lo que puede reflejar la interacción existente entre la velocidad y la memoria, así como su efecto en el procesamiento cognitivo; o quizá el papel que desempeña la memoria verbal en el rendimiento de la velocidad de procesamiento en estas edades de la infancia. Todas las pruebas

de Velocidad de procesamiento muestran correlaciones moderadamente altas con la prueba Velocidad de denominación, lo cual se esperaba dado que en todas ellas es fundamental la velocidad.

En cuanto a los índices secundarios, el INV y el ICG de la WPPSI-IV muestran un patrón de relaciones con la NEPSY-II similar al observado con el CI total. De estos índices, el ICC muestra la correlación más alta con las pruebas de la NEPSY-II debido a la correlación elevada del IMT y el IVP de la WPPSI-IV con las pruebas de la NEPSY-II (p. ej., *Memoria narrativa, Procesamiento fonológico, Velocidad de denominación* e *Inhibición*). Estos resultados aportan más datos acerca de la validez de estas nuevas puntuaciones compuestas.

En general, las observaciones realizadas en este estudio fueron las esperadas. Los resultados de este estudio respaldan la validez convergente y discriminante de las puntuaciones de las pruebas y de las puntuaciones compuestas de la WPPSI-IV.

Correlaciones con el BASC-2 CPP

El Cuestionario para padres del BASC-2 (BASC-2 CPP) es una escala que se basa en descripciones de conductas observables en un niño y proporciona información de la conducta y personalidad del niño. Forma parte del BASC-2, un sistema multimétodo y multidimensional de evaluación para las edades de 2 años a 25 años 11 meses, que se utiliza en la realización de diagnósticos diferenciales y en la clasificación de trastornos emocionales y de la conducta.

En este estudio comparativo con la WPPSI-IV se seleccionaron tres escalas del BASC-2 CPP con el objetivo de obtener más información sobre las relaciones entre la aptitud cognitiva y las dimensiones conductuales medidas por las escalas seleccionadas. Las escalas de la BASC-2 CPP utilizadas fueron: Problemas de atención (tendencia a la distracción y falta de capacidad para mantener la concentración), Funcionamiento ejecutivo (aptitud para controlar la conducta mediante la planificación, la anticipación, la inhibición o la realización de una actividad dirigida, de forma que la reacción sea adecuada al entorno), y Autocontrol emocional (regulación de los afectos y emociones en las reacciones al entorno). Las escalas se miden mediante puntuaciones T, con una media 50 y una desviación típica 10; las puntuaciones de 60 a 69 indican que se da una situación de riesgo, y las puntuaciones iguales o superiores a 70 se clasifican como clínicamente significativas. Aunque el funcionamiento ejecutivo y el autocontrol emocional pueden parecer rasgos positivos, las puntuaciones altas en las escalas correspondientes indican que el funcionamiento ejecutivo y el autocontrol emocional son insuficientes, respectivamente.

Se esperaba que, en general, las correlaciones inversas ente la WPPSI-IV y determinadas puntuaciones del BASC-2 CPP fueran bajas. Respecto a los índices primarios de la WPPSI-IV, se predijo que el IMT sería el que mostraría la correlación más alta con la escala Problemas de atención del BASC-2 CPP, puesto que comparten parte del constructo en algunos modelos de la memoria de trabajo (Cowan, 1988). También se esperaba que el ICC mostrara correlaciones más altas con las escalas Problemas de atención y Funcionamiento ejecutivo que el ICG, dado que el ICC se basa en procesos cognitivos que a menudo se clasifican como funciones ejecutivas y porque la atención se incluye en el funcionamiento ejecutivo.

Se aplicó la WPPSI-IV y determinadas escalas del BASC-2 CPP, en un único día, a 1984 niños no clínicos de 2 años 6 meses a 7 años 6 meses, así como a sus cuidadores. La tabla 6.13 muestra las medias, las desviaciones típicas y los coeficientes de correlación corregidos.

Tabla 6.13. Correlaciones entre la WPPSI-IV y el BASC-2 CPP

	Escala del BASC-2 CPP			WPPSI-IV		
Prueba/punt. de procesamiento/ punt. compuesta de la WPPSI-IV	**Problemas de atención**	**Funcionamiento ejecutivo**	**Autocontrol emocional**	**Media**	***Dt***	***n***
Información	–.17	–.15	–.10	10.5	2.6	1,983
Semejanzas	–.14	–.14	–.11	10.3	2.7	1,371
Vocabulario	–.18	–.17	–.12	10.3	2.7	1,374
Comprensión	–.13	–.12	–.09	10.3	2.8	1,372
Dibujos	–.15	–.14	–.10	10.5	2.8	1,984
Nombres	–.17	–.13	–.05	10.5	2.8	1,981
Cubos	–.11	–.12	–.09	10.2	2.8	1,968
Rompecabezas	–.13	–.11	–.07	10.3	2.8	1,982
Matrices	–.14	–.11	–.06	10.3	2.8	1,375
Conceptos	–.14	–.14	–.11	10.3	2.9	1,375
Reconocimiento	–.15	–.13	–.09	10.2	2.8	1,982
Localización	–.18	–.15	–.11	10.2	2.8	1,983
Búsqueda de animales	–.22	–.17	–.13	10.2	2.8	1,366
Cancelación	–.10	–.10	–.09	10.2	2.8	1,373
Clave de figuras	–.18	–.13	–.09	10.3	2.9	1,373
Cancelación aleatoria	–.08	–.10	–.08	10.1	2.9	1,373
Cancelación estructurada	–.11	–.09	–.05	10.2	2.8	1,373
ICV	–.18	–.17	–.11	102.2	13.3	1,979
IVE	–.14	–.14	–.10	101.7	13.6	1,966
IRF	–.18	–.17	–.11	101.9	13.6	1,375
IMT	–.22	–.17	–.13	101.2	13.8	1,981
IVP	–.19	–.17	–.13	101.2	13.5	1,364
CIT	–.24	–.22	–.15	102.2	13.1	1,952
IAV	–.19	–.17	–.09	102.5	13.5	1,981
INV	–.25	–.22	–.16	101.9	13.3	1,956
ICG	–.21	–.19	–.13	102.2	13.1	1,963
ICC	–.24	–.21	–.16	101.2	12.9	1,362
BASC-2 CPP						
Media	49.4	48.3	49.9			
Dt	8.3	8.8	9.6			
n	1,984	1,981	1,983			

Nota: Se corrigieron las correlaciones de acuerdo con la variabilidad de la muestra de tipificación estadounidense de la WPPSI-IV (Guilford & Fruchter, 1978).

Las medias de las puntuaciones compuestas de la WPPSI-IV y la media de las puntuaciones del BASC-2 CPP se sitúan en el rango promedio. Las medias de las puntuaciones de las pruebas de la WPPSI-IV también se sitúan cerca de la media.

Como se esperaba, las correlaciones entre la WPPSI-IV y determinadas puntuaciones del BASC-2 CPP son negativas y bajas. Estos resultados indican que los niños de una población no clínica muestran muy pocos síntomas de problemas conductuales. Aunque las correlaciones son estadísticamente significativas, son bajas, lo que indica que la asociación entre el rendimiento cognitivo y la conducta es mínima.

En cuanto a los índices primarios de la WPPSI-IV, tal y como se esperaba, el IMT muestra las correlaciones más altas con la escala Problemas de atención del BASC-2 CPP; no obstante, las correlaciones también son bajas. Como se predijo, el ICC muestra una correlación ligeramente más alta con las escalas Problemas de atención y Funcionamiento ejecutivo que el ICG. El CI total muestra la mayor correlación con Problemas de atención.

Respecto a las pruebas de la WPPSI-IV, *Búsqueda de animales, Vocabulario, Localización* y *Clave de figuras* muestran las correlaciones más altas con Problemas de atención. Además, *Vocabulario* y *Búsqueda de animales* muestran correlaciones más altas con Funcionamiento ejecutivo que con otras pruebas de la WPPSI-IV. Y aunque *Búsqueda de animales* muestra una correlación alta con la escala Autocontrol emocional, esta correlación es casi insignificante. Estos resultados sugieren, por una parte, que las aptitudes cognitivas muestran poca asociación con la atención y el funcionamiento ejecutivo, evaluados con una medida basada en el informe de los padres; y, por otra, que tales aptitudes no están relacionadas significativamente con el autocontrol emocional en esta muestra no clínica.

Grupos especiales

Los resultados de los estudios realizados con grupos especiales proporcionan evidencias adicionales de validez de criterio. Las evidencias de la validez de una escala, aplicada a grupos clínicos, es fundamental cuando sus resultados forman parte de una evaluación diagnóstica exhaustiva. Durante la fase de de tipificación estadounidense de la WPPSI-IV se llevaron a cabo varios estudios con grupos especiales a fin de determinar si los constructos medidos por la escala resultaban como se esperaba en determinados grupos con características definidas.

Un equipo de examinadores e investigadores independientes recopilaron los datos para los estudios con grupos especiales. Los sujetos que participaron en estos estudios se seleccionaron en distintos ámbitos clínicos y fueron aceptados para participar en la muestra de grupos especiales de acuerdo con ciertos criterios de inclusión específicos (en el anexo D de este manual figuran los criterios seguidos para cada uno de los grupos especiales).

Es importante señalar las limitaciones de estos estudios. En primer lugar, cabe destacar que las muestras no se seleccionaron aleatoriamente, sino en función de la disponibilidad. Por ello, es posible que estos estudios no sean representativos de la categoría diagnóstica en su totalidad. En segundo lugar, dado que los datos de cada uno de los grupos especiales se recogieron en diferentes ámbitos clínicos, los diagnósticos de los niños del mismo grupo pueden haberse basado en criterios y procedimientos diferentes. Además, el tamaño de las muestras de algunos de estos grupos es pequeño y cubre solo una parte del rango de edad de la WPPSI-IV. Finalmente, solo se describe el rendimiento del grupo. Por todo ello, los datos de estas muestras se presentan a modo de ejemplo, y no pretenden ser totalmente representativos de los grupos diagnósticos. El objetivo de estos estudios es demostrar que la WPPSI-IV puede ofrecer estimaciones válidas de la aptitud intelectual de los niños de estos grupos especiales, y que los niños que presentan trastornos específicos obtienen las puntuaciones esperadas. Sin embargo, las puntuaciones de la WPPSI-IV no deberían usarse nunca como el único criterio para establecer un diagnóstico o una clasificación.

Las características demográficas de las muestras de los grupos especiales de la WPPSI-IV se presentan en la tabla 6.14. Todos los grupos control, seleccionados aleatoriamente de la muestra de tipificación, se clasificaron en función de estas variables demográficas. Las tablas 6.15-6.26 presentan las medias de las puntuaciones de la WPPSI-IV para los grupos especiales y para sus correspondientes grupos control.

UAM
BIBLIOTECA DE PSICOLOGÍA

Tabla 6.14. Datos demográficos de los grupos especiales (datos estadounidenses)

		Discapacidad intelectual										
	Alta capacidad intelectual	Leve	Moderada	Retraso del desarrollo cognitivo	Factores de riesgo para el desarrollo	Problemas previos a la lectoescritura	TDAH	Comportamiento perturbador	Trastorno del lenguaje expresivo	Trastorno del lenguaje mixto receptivo-expresivo	Trastorno autista	Trastorno de Asperger
N	56	39	34	72	42	37	53	26	25	42	38	38
Edad												
Media	6.2	5.3	5.3	4.4	5.3	6.2	5.9	6.2	5.6	5.5	5.5	6.2
Dt	0.9	1.5	1.2	1.5	1.3	0.7	1.2	0.7	1.1	0.9	1.2	0.9
Sexo												
Mujer	51.8	33.3	38.2	36.1	35.7	54.1	24.5	7.7	52.0	35.7	21.1	21.1
Varón	48.2	66.7	61.8	63.9	64.3	45.9	75.5	92.3	48.0	64.3	78.9	78.9
Nivel educativo de los padres												
≤8 años	–	10.3	2.9	–	–	–	–	–	–	–	–	–
9-11 años	–	25.6	11.8	25.0	4.8	8.1	15.1	15.4	4.0	14.3	10.5	2.6
12 años	1.8	25.6	41.2	31.9	14.3	40.5	22.6	46.2	16.0	31.0	26.3	26.3
13-15 años	7.1	17.9	20.6	23.6	31.0	35.1	34.0	26.9	36.0	31.0	39.5	18.4
≤16 años	91.1	20.5	23.5	19.4	50.0	16.2	28.3	11.5	44.0	23.8	23.7	52.6
Zona geográfica												
Medio Oeste	26.8	38.5	41.2	27.8	33.3	21.6	32.1	26.9	28.0	26.2	23.7	15.8
Noreste	8.9	5.1	2.9	11.1	7.1	24.3	13.2	7.7	12.0	11.9	10.5	10.5
Sur	48.2	38.5	52.9	30.6	31.0	21.6	43.4	42.3	24.0	28.6	47.4	39.5
Oeste	16.1	17.9	2.9	30.6	28.6	32.4	11.3	23.1	36.0	33.3	18.4	34.2

Nota: Excepto para el tamaño de la muestra (*N*) y la edad, los datos se muestran en porcentajes. Si el porcentaje total no suma 100, es debido al redondeo.

Niños con alta capacidad intelectual

Los niños con alta capacidad intelectual muestran una gran aptitud en el funcionamiento intelectual, la flexibilidad cognitiva, la creatividad y/u otras áreas específicas (Geake, 2008; Koziol, Budding y Chidekel, 2010). Los niños con alta capacidad intelectual no se caracterizan simplemente por pensar como adultos; al igual que el resto de los niños, experimentan cambios en el pensamiento conceptual relacionados con la edad, de acuerdo con los modelos neopiagetianos. Sin embargo, estos niños obtienen un gran beneficio de las experiencias que son importantes para desarrollar aptitudes en áreas específicas (Loewen, 2006). Diversos estudios han analizado el rendimiento de niños con alta capacidad intelectual en medidas de la aptitud cognitiva. Los resultados sugieren que las medias de las puntuaciones de estos niños son significativamente más altas que las de los niños de la población general (Rimm et al., 2008; Rowe, Kingsley y Thompson, 2010; Sweetland, Reina y Tatti, 2006; Wechsler, 2002, 2003).

Habitualmente las personas con alta capacidad intelectual muestran un buen rendimiento tanto en las pruebas basadas en la aptitud verbal como en las pruebas no verbales de las escalas de inteligencia de Wechsler. Sin embargo, en algunas ocasiones poco frecuentes, pueden mostrar grandes divergencias entre las puntuaciones obtenidas en pruebas verbales y pruebas no verbales (Sweetland et al., 2006). Normalmente, los puntos fuertes de los niños con alta capacidad intelectual se reflejan en las áreas de comprensión verbal, aptitud visoespacial y razonamiento fluido; sin embargo, según la mayoría de los estudios, muestran un rendimiento más bajo en memoria de trabajo y velocidad de procesamiento (Raiford et al., 2005; Rimm et al., 2008; Rowe et al., 2010). No obstante, su rendimiento en tareas de memoria de trabajo y de velocidad de procesamiento es generalmente más alto que el de la población general (véanse las excepciones observadas en el área de la velocidad de procesamiento en Rimm et al., 2008 y Rowe et al., 2010). La relativamente alta memoria de trabajo de estos niños parece favorecer la creatividad, que es un aspecto importante de la alta capacidad intelectual (Geake, 2008).

Puesto que está demostrado que los niños con alta capacidad intelectual muestran puntos débiles intraindividuales en las áreas de la memoria de trabajo y la velocidad de procesamiento, algunos autores recomiendan utilizar el CI total y el ICG cuando se usan las puntuaciones de la aptitud cognitiva como criterio para identificar la alta capacidad intelectual en estos niños (p. ej., para su admisión en un programa especializado) (Rimm et al., 2008). No obstante, el CI total parece ser un mejor indicador del rendimiento académico de los niños con alta capacidad intelectual; y, de entre los índices de la WISC-IV, el ICV y el IMT son indicadores sólidos del rendimiento académico de estos niños (Rowe et al., 2010). Por ello, el objetivo de la evaluación (p. ej., admisión en un programa especializado) puede determinar qué puntuaciones van a utilizarse.

La interpretación de los resultados de estudios centrados en la identificación de la alta capacidad intelectual en los niños pequeños es complicada por diversos factores. En primer lugar, las distintas medidas de aptitud cognitiva aportan estimaciones altamente correlacionadas, pero diferentes. En segundo lugar, los criterios para identificar a los niños con alta capacidad intelectual varían de un lugar a otro. En tercer lugar, los niños pequeños identificados como candidatos han sido evaluados recientemente, puesto que acaban de iniciar su escolarización. Esto puede dar lugar a que sean eliminados del estudio, puesto que el hecho de haber sido evaluados recientemente puede producir los efectos de aplicación repetida y, por consiguiente, puntuaciones infladas. En cuarto lugar, los niños que participan en programas especializados en alta capacidad intelectual son admitidos, a veces, en función de las puntuaciones que obtienen en medidas del rendimiento académico, y no siempre las puntuaciones altas en estas medidas se corresponden con una aptitud cognitiva alta.

Tomando como referencia las investigaciones previas, se esperaba que el grupo de niños con alta capacidad intelectual obtuviera puntuaciones medias significativamente superiores a las del grupo control en todas las puntuaciones de las pruebas y las puntuaciones compuestas de la WPPSI-IV; y, dado que el IVP y el ICC dependen de la velocidad de procesamiento, las medias de estos índices serían relativamente inferiores a las medias del resto de las puntuaciones compuestas.

Se aplicó la WPPSI-IV a 56 niños de 4:2 a 7:6 años, que habían sido previamente identificados como niños con alta capacidad intelectual. Para poder formar parte de este estudio, los niños tenían que haber obtenido una puntuación por encima de 2 desviaciones típicas en alguna prueba estandarizada de aptitud intelectual general. También se aceptó a niños que participaban en algún programa de alta capacidad intelectual aunque no se tuvieran sus puntuaciones. La tabla 6.15 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con alta capacidad intelectual y para el grupo control pareado.

Tabla 6.15. Resultados del grupo con alta capacidad intelectual y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ puntuación compuesta	Alta capacidad intelectual		Grupo control		Diferencia entre medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	14.8	2.5	11.1	2.1	56	−3.64	−7.61	<.01	−1.58
Semejanzas	14.9	2.4	11.2	2.5	55	−3.73	−7.29	<.01	−1.52
Vocabulario	15.3	2.2	11.3	2.9	55	−4.00	−8.74	<.01	−1.55
Comprensión	14.5	2.6	11.0	2.3	56	−3.48	−7.62	<.01	−1.42
Dibujos	14.4	2.8	11.1	2.6	56	−3.30	−6.21	<.01	−1.22
Nombres	14.3	2.4	11.6	2.7	56	−2.64	−5.33	<.01	−1.03
Cubos	14.3	3.1	10.7	2.9	56	−3.54	−6.47	<.01	−1.18
Rompecabezas	12.6	2.7	11.1	2.3	56	−1.50	−3.55	<.01	−0.60
Matrices	14.5	2.5	10.8	2.8	56	−3.68	−7.95	<.01	−1.39
Conceptos	12.8	2.1	11.0	2.4	55	−1.82	−4.28	<.01	−0.81
Reconocimiento	13.2	2.8	10.6	3.1	55	−2.58	−4.20	<.01	−0.87
Localización	13.0	2.9	10.2	2.7	56	−2.84	−5.22	<.01	−1.01
Búsqueda de animales	11.9	3.3	10.6	2.5	56	−1.27	−2.11	.04	−0.43
Cancelación	11.4	3.1	10.0	2.5	55	−1.40	−2.64	.01	−0.50
Clave de figuras	11.4	2.2	10.6	3.3	56	−0.73	−1.40	.17	−0.26
Cancelación aleatoria	11.3	3.2	10.1	2.6	55	−1.20	−2.03	.05	−0.41
Cancelación estructurada	11.2	3.1	9.9	2.8	55	−1.29	−2.54	.01	−0.44
ICV	128.1	12.8	106.0	11.9	55	−22.11	−8.58	<.01	−1.79
IVE	120.2	14.6	105.4	11.2	56	−14.86	−6.26	<.01	−1.14
IRF	121.5	10.2	105.5	12.6	55	−15.98	−8.25	<.01	−1.39
IMT	118.4	13.9	102.7	13.6	55	−15.73	−5.59	<.01	−1.14
IVP	110.2	15.2	102.0	12.3	55	−8.16	−3.01	<.01	−0.59
CIT	127.2	9.4	105.8	12.5	54	−21.43	−9.95	<.01	−1.94
IAV	124.1	12.9	107.6	12.4	56	−16.50	−6.36	<.01	−1.30
INV	123.2	8.9	105.4	13.2	54	−17.76	−8.76	<.01	−1.58
ICG	129.3	10.2	106.0	12.1	55	−23.29	−10.63	<.01	−2.08
ICC	117.3	11.6	102.7	12.7	54	−14.56	−6.25	<.01	−1.20

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Como se esperaba, las medias de las puntuaciones compuestas obtenidas por el grupo con alta capacidad intelectual son significativamente superiores a las del grupo control pareado. Las medias de los índices primarios del grupo con alta capacidad intelectual oscilan entre 110.2 (IVP) y 128.1 (ICV), y la media del CI total es de 127.2. Los tamaños del efecto para las diferencias entre las puntuaciones compuestas son grandes, excepto para el IVP, cuyo tamaño del efecto es moderado. La media del IRF es más alta que la del IVE, y el tamaño del efecto también es algo más grande. Los tamaños del efecto del IVE y del IMT para las diferencias entre las medias son idénticos. La media del IVP fue relativamente más baja que las medias del resto de las puntuaciones compuestas; de hecho, las medias más bajas en el grupo de alta capacidad intelectual se obtuvieron en las pruebas *Cancelación* y *Clave de figuras*. Las puntuaciones de los índices secundarios oscilan entre 117.3 (ICC) y 129.3 (ICG). El tamaño del efecto del INV es más pequeño que el del CI total o del ICG.

Todas las medias de las puntuaciones escalares de las pruebas son significativamente más altas que las del grupo control pareado, excepto la de *Clave de figuras*, que tiene un tamaño del efecto pequeño. En cuanto a las pruebas, los tamaños del efecto más grandes para las diferencias entre las medias de las puntuaciones escalares se obtuvieron en *Información, Vocabulario* y *Semejanzas*. Los resultados de las pruebas son consistentes con los obtenidos en estudios previos (Wechsler, 2002, 2003, 2008).

Otros análisis realizados con la WPPSI-IV indican que de los niños con alta capacidad intelectual, el 85 % obtienen un CI total igual o superior a 120, el 84 % obtienen un ICG igual o superior a 120 y el 67 % obtiene un INV igual o superior a 120. En cambio, solo el 20 %, el 11 % y el 16 % de los niños del grupo control obtuvieron las respectivas puntuaciones.

Las puntuaciones más altas del grupo de niños con alta capacidad intelectual confirman su identificación como tales, y siguen la misma línea de los estudios comparativos realizados (Rimm et al., 2008; Rowe et al., 2010; Wechsler, 2002, 2003, 2008). Al igual que este estudio, diversas investigaciones anteriores también indican que los niños con alta capacidad intelectual muestran un rendimiento más bajo en las pruebas diseñadas para medir la velocidad de procesamiento que en las pruebas de otros dominios cognitivos (Rimm et al., 2008; Rowe et al., 2010; Wechsler, 2002, 2003). El IRF muestra una media algo más alta y un tamaño del efecto más grande que el IVE, lo que sugiere que ambos índices pueden tener asociaciones diferenciales con la alta capacidad intelectual. Esto puede reflejar la fuerte relación entre el razonamiento fluido y la inteligencia general. El menor tamaño del efecto del INV respecto al del CI total o al del ICG puede indicar que la comprensión verbal es un componente importante de los métodos generales de evaluación de la alta capacidad intelectual, en concreto en entornos escolares basados en los contenidos verbales. Los resultados de la WPPSI-IV y los obtenidos por grupos con alta capacidad intelectual evaluados con otras escalas de inteligencia de Wechsler previas también indican que estos instrumentos miden constructos similares. Estos resultados demuestran que la WPPSI-IV es una herramienta útil para la detección de la alta capacidad intelectual. Este estudio se llevó a cabo con niños a los que ya se les había diagnosticado alta capacidad intelectual; por lo que los resultados no deberían generalizarse a niños que presenten un alto rendimiento en otros dominios. Se requieren más investigaciones para evaluar los resultados obtenidos en la WPPSI-IV por parte de niños con alto rendimiento en otros dominios.

Niños con discapacidad intelectual leve o moderada

De acuerdo con los criterios especificados en el DSM-IV-TR (American Psychiatric Association, 2000, 2002), un sujeto diagnosticado con discapacidad intelectual leve debe mostrar «un funcionamiento intelectual significativamente inferior al promedio» y limitaciones significativas en el funcionamiento adaptativo en dos o más de las siguientes áreas: comunicación, cuidado de sí mismo, vida doméstica, habilidades sociales/interpersonales, utilización de recursos comunitarios, autocontrol, habilidades académicas funcionales, trabajo, ocio, y salud y seguridad. De acuerdo con la AAIDD, la discapacidad intelectual se origina antes de los 18 años y se caracteriza por una serie de limitaciones significativas tanto en el funcionamiento intelectual como en el comportamiento adaptativo expresado en las habilidades prácticas, conceptuales y sociales (AAIDD, 2010). Tanto el DSM-IV-TR como la AAIDD coinciden al señalar que un rendi-

miento significativamente bajo en un test de aptitud intelectual general, como la WPPSI-IV, es un criterio necesario pero no suficiente para diagnosticar discapacidad intelectual; también hay que evaluar el funcionamiento adaptativo.

Asimismo destacan la necesidad de considerar el error típico de medida de la escala al interpretar las puntuaciones, haciendo hincapié en la importancia de tener en cuenta distintos criterios y de aplicar cuidadosamente el juicio clínico en la identificación de niños con discapacidad intelectual. En una distribución normal de puntuaciones CI, alrededor del 2.2 % de sujetos obtienen puntuaciones como mínimo de 2 desviaciones típicas por debajo de la media (p. ej., CI ≤ 70). La prevalencia de la discapacidad intelectual varía de un estudio a otro. Un metaanálisis exhaustivo, basado en 52 estudios, indicó que la prevalencia de la discapacidad intelectual en la población general era de 10.37 por 1000 (Maulik, Mascarenhas, Mathers, Dua y Saxena, 2011), o aproximadamente del 1 %, si se tiene en cuenta que el diagnóstico de la discapacidad intelectual debe establecer los niveles apropiados tanto de la aptitud intelectual como del funcionamiento adaptativo.

Se han llevado a cabo numerosos estudios para evaluar el rendimiento de los sujetos con discapacidad intelectual en versiones anteriores de las escalas de inteligencia de Wechsler. Al respecto se ha demostrado que la prevalencia de diferencias marcadas e inusuales entre los índices verbales y los no verbales disminuye al mismo tiempo que disminuye el nivel de aptitud (Gordon, Duff, Davidson y Whitaker, 2010; Spruill, 1998; Wechsler, 2002, 2003). En algunos estudios se ha observado que el rendimiento en comprensión verbal y en memoria de trabajo era relativamente más bajo comparado con el rendimiento en razonamiento perceptivo y en velocidad de procesamiento (Gordon et al., 2010). Además, las desviaciones típicas de las pruebas y de las puntuaciones compuestas son menores para los sujetos con discapacidad intelectual que para los sujetos de la población general (Wechsler, 2002, 2003, 2008). Así pues, parece que la variabilidad en el rendimiento, tanto en las pruebas como en las puntuaciones compuestas, es menor en los sujetos con discapacidad intelectual que en los sujetos de la población general.

En el estudio de la WPPSI-IV, se esperaba que las medias de las puntuaciones compuestas obtenidas por los niños con discapacidad intelectual se situaran entre 2 y 3 desviaciones típicas por debajo de las medias obtenidas por los niños del grupo control pareado. Se supuso también que los niños con discapacidad intelectual presentarían menos variabilidad (es decir, menor desviación típica) en las pruebas y las puntuaciones compuestas en comparación con el grupo control pareado. Se aplicó la WPPSI-IV a 73 niños no institucionalizados de 2:6 a 7:6 años, que habían sido diagnosticados con discapacidad intelectual. El grupo estaba formado por 39 niños con discapacidad intelectual leve y 34 con discapacidad intelectual moderada. En pruebas estandarizadas de aptitud cognitiva, los niños del grupo con discapacidad intelectual leve habían obtenido puntuaciones de 2-3 desviaciones típicas por debajo de la media (p. ej., 55 ≤ CI total ≤ 70); y los niños del grupo de discapacidad intelectual moderada, de 3-4 desviaciones típicas por debajo de la media (p. ej., 40 ≤ CI total ≤ 54). Las tablas 6.16 y 6.17 presentan las medias de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas obtenidas por niños con discapacidad intelectual leve y moderada y por los grupos control pareados.

Tabla 6.16. Resultados del grupo con discapacidad intelectual leve y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ puntuación compuesta	Discapacidad intelectual leve		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	3.7	2.2	10.2	2.9	39	6.49	10.27	<.01	2.52
Semejanzas	3.0	1.9	9.5	3.4	31	6.55	9.52	<.01	2.38
Vocabulario	3.9	1.7	9.7	3.2	31	5.87	8.70	<.01	2.29
Comprensión	4.3	1.9	9.5	3.2	31	5.19	8.12	<.01	1.97
Dibujos	4.4	2.7	9.6	2.2	39	5.23	10.03	<.01	2.12
Nombres	3.9	1.6	9.8	3.1	39	5.90	10.66	<.01	2.39
Cubos	3.9	2.0	9.9	2.2	38	6.00	13.21	<.01	2.85
Rompecabezas	3.7	1.9	9.5	2.6	39	5.79	11.85	<.01	2.54
Matrices	4.4	1.9	8.7	2.5	31	4.29	8.27	<.01	1.93
Conceptos	4.5	1.8	9.7	2.9	31	5.23	9.82	<.01	2.17
Reconocimiento	4.8	2.0	10.4	2.9	39	5.59	9.53	<.01	2.24
Localización	5.4	2.5	9.7	2.1	39	4.31	8.96	<.01	1.87
Búsqueda de animales	4.1	2.4	8.9	2.7	29	4.86	6.64	<.01	1.90
Cancelación	4.5	2.8	9.1	3.5	31	4.52	5.68	<.01	1.43
Clave de figuras	4.2	2.6	9.2	2.6	31	5.00	7.42	<.01	1.92
Cancelación aleatoria	4.8	2.7	9.0	3.7	31	4.16	5.11	<.01	1.28
Cancelación estructurada	4.9	3.0	9.0	3.2	31	4.13	5.47	<.01	1.33
ICV	66.1	8.7	99.4	15.0	39	33.26	11.65	<.01	2.71
IVE	65.1	9.0	97.9	11.6	38	32.76	15.26	<.01	3.16
IRF	68.7	8.3	95.3	13.3	31	26.65	10.20	<.01	2.40
IMT	72.1	10.8	100.3	12.8	39	28.21	11.10	<.01	2.38
IVP	69.7	11.2	94.2	14.3	29	24.45	7.31	<.01	1.90
CIT	63.6	5.5	97.4	13.7	36	33.81	13.10	<.01	3.24
IAV	68.5	9.3	98.2	12.7	39	29.72	11.99	<.01	2.67
INV	64.8	6.1	96.8	12.9	36	31.97	14.02	<.01	3.17
ICG	62.6	6.3	97.0	13.2	38	34.39	13.81	<.01	3.33
ICC	66.1	10.2	96.5	13.7	29	30.38	9.86	<.01	2.52

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Las medias de los índices primarios del grupo con discapacidad intelectual leve oscilan entre 65.1 (IVE) y 72.1 (IMT), y la media del CI total es de 63.6. Tanto la media del CI total como las de los índices primarios son significativamente más bajas que las medias correspondientes del grupo control pareado. Como se esperaba, las medias de los índices primarios obtenidas por el grupo con discapacidad intelectual moderada son incluso más bajas, entre 53.3 (IVP) y 56.9 (IRF). La media del CI total de los niños con discapacidad intelectual moderada es de 50.0 y los tamaños del efecto de las diferencias entre grupos para las medias de los índices primarios y el CI total son grandes.

Las medias de los índices secundarios del grupo con discapacidad intelectual leve son de 62.6 (ICG) a 68.5 (IAV), y concuerdan con los valores esperados. La media del INV obtenida por este grupo es de 64.8, un valor muy similar al del CI total, aunque ligeramente superior a este. El tamaño del efecto del INV, aun siendo grande, es ligeramente más pequeño que el del CI total. Las medias de los índices secundarios del grupo con discapacidad intelectual moderada oscilan entre 47.6 (ICC) y 55.0 (IAV), y también concuerdan con los valores esperados. La media del INV obtenida por este grupo es de 51.8, un valor ligeramente superior al del CI total. El tamaño del efecto del INV, aun siendo grande, es ligeramente más pequeño que el del CI total. Los resultados del INV, analizados en conjunto y comparados con los del CI total, pueden indicar

Tabla 6.17. Resultados del grupo con discapacidad intelectual moderada y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ puntuación compuesta	Discapacidad *intelectual moderada*		*Grupo control*		*Diferencia de medias de los grupos*				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	1.5	1.2	10.2	2.7	33	8.67	17.22	<.01	4.15
Semejanzas	2.1	1.3	9.8	2.7	27	7.67	12.45	<.01	3.62
Vocabulario	2.1	1.2	10.2	2.7	26	8.12	13.51	<.01	3.89
Comprensión	2.1	1.1	9.7	2.7	27	7.56	13.60	<.01	3.67
Dibujos	1.8	1.3	10.0	2.6	34	8.24	17.44	<.01	4.01
Nombres	2.2	1.3	10.4	3.4	33	8.12	12.28	<.01	3.15
Cubos	1.8	1.6	10.2	3.4	33	8.42	14.47	<.01	3.17
Rompecabezas	2.7	1.8	9.6	2.2	34	6.82	14.54	<.01	3.39
Matrices	2.6	1.6	9.6	2.9	27	6.96	10.17	<.01	2.97
Conceptos	2.4	1.8	9.5	3.3	27	7.11	8.89	<.01	2.67
Reconocimiento	2.2	1.7	10.3	2.9	34	8.09	13.88	<.01	3.40
Localización	2.4	1.9	9.4	2.7	34	6.97	12.53	<.01	2.99
Búsqueda de animales	2.0	2.2	8.4	3.0	27	6.37	8.50	<.01	2.42
Cancelación	2.1	2.3	10.4	3.0	27	8.30	15.33	<.01	3.11
Clave de figuras	3.1	2.1	9.7	3.1	26	6.65	8.50	<.01	2.51
Cancelación aleatoria	2.3	2.5	10.3	2.6	27	7.96	14.39	<.01	3.12
Cancelación estructurada	2.3	2.4	10.4	3.2	27	8.07	14.70	<.01	2.85
ICV	53.8	9.0	100.0	13.0	33	46.27	15.28	<.01	4.14
IVE	54.9	10.1	100.2	13.7	33	45.27	15.58	<.01	3.76
IRF	56.9	9.8	97.4	15.7	27	40.44	10.44	<.01	3.09
IMT	54.8	10.7	98.9	13.2	34	44.15	15.52	<.01	3.67
IVP	53.3	12.2	96.7	13.1	27	43.48	16.74	<.01	3.43
CIT	50.0	9.0	99.6	15.0	32	49.53	15.67	<.01	4.00
IAV	55.0	8.9	100.7	13.7	33	45.70	15.65	<.01	3.96
INV	51.8	9.2	98.6	15.6	33	46.79	14.65	<.01	3.65
ICG	49.8	8.7	100.3	14.1	32	50.56	16.80	<.01	4.32
ICC	47.6	8.5	96.9	13.7	27	49.30	17.54	<.01	4.32

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

que la comprensión verbal es un componente importante de la inteligencia general, como ya se ha visto respecto al grupo con alta capacidad intelectual.

La variabilidad en el rendimiento de los grupos con discapacidad intelectual leve y moderada es pequeña. Las desviaciones típicas de los índices primarios y del CI total oscilaron entre 5.5 (CI total) y 11.2 (IVP) en el grupo con discapacidad intelectual leve, y entre 9.0 (ICV y CI total) y 12.2 (IVP) en el grupo con discapacidad intelectual moderada. Ambos rangos son mucho *más pequeños que la desviación típica observada en la población general (15). En cuanto a las* desviaciones típicas de las pruebas se observa un patrón similar.

Ambos grupos obtuvieron, en todas las pruebas, puntuaciones significativamente inferiores a las correspondientes del grupo control pareado. En el grupo con discapacidad intelectual leve, los tamaños del efecto más grandes para las diferencias entre las medias de las puntuaciones escalares se obtuvieron en *Cubos, Rompecabezas, Información, Nombres, Semejanzas, Vocabulario* y *Reconocimiento*. En el grupo con discapacidad intelectual moderada, los tamaños del efecto más grandes para las diferencias entre las medias de las puntuaciones escalares se obtuvieron en *Información, Dibujos, Vocabulario, Comprensión, Semejanzas, Reconocimiento, Rompecabezas, Cubos* y *Nombres*. Estos puntos débiles relativos son consistentes con los resultados de los estudios que indicaban que el rendimiento de los niños con discapacidad intelectual era bajo en las pruebas de Comprensión verbal (Wechsler, 2002, 2003), lo que refleja la estrecha relación entre la aptitud verbal y

la inteligencia. Los resultados obtenidos también son consistentes con los estudios que señalan puntos fuertes relativos en las áreas de la memoria de trabajo (Baddeley y Jarrold, 2007; Conners, Rosenquist, Arnett, Moore y Hume, 2008; Van der Molen, Van Luit, Jongmans y Van der Molen, 2007) y la aptitud visoespacial (Caffrey y Fuchs, 2007; Fontana, 2004; Wechsler, 2003).

Del grupo con discapacidad intelectual leve, un 97 % de los niños obtuvieron un CI total y un INV iguales o inferiores a 75, frente a un 3 % y a un 5 %, respectivamente de los niños del grupo control pareado. En cambio, del grupo con discapacidad intelectual moderada, un 81 % y un 85% de los niños obtuvieron un CI total y un INV, respectivamente, iguales o inferiores a 60, pero ningún niño del grupo control obtuvo puntuaciones en ese rango. Los resultados de estos grupos demuestran que la WPPSI-IV es una herramienta útil para la detección de la discapacidad intelectual.

Niños con retraso del desarrollo cognitivo

La ley IDEIA (2004) incluye el retraso del desarrollo como una categoría para los niños de hasta 9 años, con retrasos constatados y significativos en una o más de las siguientes áreas de desarrollo: cognitivo, adaptativo, comunicativo, social o emocional, o motor. También se utiliza la categoría de retraso del desarrollo para identificar a los niños que presentan este retraso y que requieren algún tipo de servicio acorde a su situación antes de que pueda ser diagnosticado formalmente con una discapacidad intelectual. El presente estudio incluyó solo niños con retraso del desarrollo cognitivo. Por lo tanto no formaron parte de la muestra los niños con retrasos del desarrollo en otras áreas, aunque dichos retrasos estuvieran relacionados con el desarrollo cognitivo (Brassard & Boehm, 2007; Ford et al., 2012).

Se esperaba que los niños con retraso del desarrollo cognitivo obtuvieran puntuaciones medias significativamente inferiores a las del grupo control pareado en todas las pruebas y en todas las puntuaciones compuestas de la WPPSI-IV. No se esperaba que estas diferencias fueran tan grandes como las observadas entre el grupo de niños con discapacidad intelectual leve y el grupo control pareado.

Se aplicó la WPPSI-IV a 72 niños de 2:6 a 7:6 años, que habían sido diagnosticados con retraso del desarrollo cognitivo. La tabla 6.18 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con retraso del desarrollo cognitivo y para el grupo control pareado.

Tabla 6.18. Resultados del grupo con retraso del desarrollo cognitivo y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ puntuación compuesta	Retraso del desarrollo cognitivo		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	6.4	2.5	9.6	3.1	71	3.17	7.05	<.01	1.13
Semejanzas	6.1	1.8	9.8	2.8	36	3.75	6.46	<.01	1.59
Vocabulario	6.1	2.2	9.3	2.8	36	3.28	5.99	<.01	1.30
Comprensión	5.4	2.3	9.3	2.9	36	3.89	6.09	<.01	1.49
Dibujos	7.0	2.1	9.5	3.1	72	2.44	6.27	<.01	0.92
Nombres	6.3	2.5	9.2	2.6	72	2.90	6.75	<.01	1.14
Cubos	7.4	2.7	9.8	2.8	69	2.38	5.11	<.01	0.87
Rompecabezas	7.5	3.1	9.3	3.2	72	1.79	3.04	<.01	0.57
Matrices	7.4	2.2	9.4	2.7	36	2.03	4.28	<.01	0.82
Conceptos	7.4	2.8	9.7	2.6	36	2.33	3.75	<.01	0.86
Reconocimiento	7.3	2.5	9.6	3.3	72	2.31	4.74	<.01	0.79
Localización	7.9	2.7	9.7	3.1	72	1.82	3.94	<.01	0.63
Búsqueda de animales	8.3	2.6	9.2	3.5	36	0.89	1.06	.30	0.29
Cancelación	7.2	3.4	9.4	3.0	35	2.14	2.57	.01	0.67
Clave de figuras	7.0	3.2	9.0	3.2	36	2.00	2.41	.02	0.63
Cancelación aleatoria	7.8	3.5	9.2	3.1	35	1.40	1.67	.10	0.42
Cancelación estructurada	*7.1*	*3.1*	*9.5*	*2.9*	*35*	*2.40*	*3.08*	*<.01*	*0.80*
ICV	81.7	9.3	98.7	15.4	71	16.94	8.46	<.01	1.33
IVE	86.4	13.8	97.7	14.9	69	11.32	4.31	<.01	0.79
IRF	84.8	11.9	97.5	13.6	36	12.69	5.02	<.01	0.99
IMT	85.8	11.9	97.8	16.6	72	12.04	5.32	<.01	0.83
IVP	87.9	14.6	96.0	18.5	35	8.14	1.84	.07	0.49
CIT	81.4	8.8	98.3	16.0	68	16.82	7.50	<.01	1.30
IAV	*81.3*	*10.2*	*96.1*	*14.2*	*72*	*14.81*	*7.88*	*<.01*	*1.20*
INV	83.1	11.6	97.9	16.3	69	14.80	6.34	<.01	1.05
ICG	81.6	9.4	98.5	14.9	68	16.85	8.01	<.01	1.35
ICC	85.9	12.7	95.4	16.4	35	9.54	2.49	.02	0.65

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Los niños con retraso del desarrollo cognitivo mostraron un rendimiento inferior al del grupo control en todas las medias de las puntuaciones compuestas, excepto en el IVP, que es significativo en el nivel $p = .10$. Las medias de los índices primarios del grupo con retraso del desarrollo cognitivo oscilan entre 81.7 (ICV) y 87.9 (IVP), y la media del CI total es de 81.4. Los tamaños del efecto para las diferencias entre las puntuaciones compuestas son grandes, excepto para el IVP y para el ICC, cuyo tamaño del efecto es pequeño y moderado, respectivamente.

Las medias de los índices secundarios oscilan entre 81.3 (IAV) y 85.9 (ICC). La media del INV es de 83.1, valor ligeramente superior al de la media del CI total, lo que probablemente indique de nuevo que la aptitud verbal es un componente importante de la inteligencia general en esta muestra; no obstante, el tamaño del efecto de la diferencia del INV entre ambos grupos es grande. Como se esperaba, las diferencias entre las medias de las puntuaciones compuestas del grupo con retraso del desarrollo cognitivo y las del grupo control pareado no son tan grandes como las observadas en el grupo con discapacidad intelectual leve y su correspondiente grupo control.

Todas las puntuaciones escalares de las pruebas obtenidas por el grupo con retraso del deterioro cognitivo son significativamente más bajas que las del grupo control, con excepción de *Búsqueda de animales*. Los tamaños del efecto más grandes para las diferencias entre las medias de las puntuaciones escalares de las pruebas se observan en *Semejanzas, Comprensión, Vocabulario, Nombres* e *Información*. Estos puntos débiles relativos son consistentes con resultados anteriores (Wechsler,

2002) y reflejan la estrecha interrelación de las destrezas lingüísticas y el desarrollo cognitivo en esas edades (Brassard y Boehm, 2007; Ford et al., 2012). Anteriormente, en esta población se habían observado puntos débiles similares en comprensión verbal (Hughes y McIntosh, 2002).

Los resultados indican que aproximadamente el 90 % de los niños con retraso del desarrollo cognitivo obtuvieron un CI total igual o inferior a 90; mientras que solo el 30 % de los niños del grupo control pareado obtuvieron un CI total igual o inferior a 90. Por otra parte, aproximadamente el 76 % de los niños con retraso del desarrollo cognitivo obtuvieron un INV igual o inferior a 90; mientras que solo el 30 % de los niños del grupo control obtuvieron un INV igual o inferior a 90.

Se requieren más investigaciones para determinar si estos patrones de resultados son consistentes entre los distintos grupos de niños con retraso del desarrollo cognitivo, dadas las distintas etiologías. Estas observaciones pueden ampliarse en nuevos estudios que se basen en niños con retraso del desarrollo en más de una área. No obstante, los resultados de este estudio demuestran que la WPPSI-IV es una herramienta útil para medir la aptitud cognitiva al evaluar exhaustivamente a niños con retraso del desarrollo cognitivo.

Niños con factores de riesgo para el desarrollo

A causa de diversos factores biológicos y ambientales, los niños tienen un mayor riesgo de presentar retrasos del desarrollo. Los factores más comunes son: bajo peso al nacer, anomalías cromosómicas, dificultad respiratoria perinatal, infección congénita, desnutrición, maltrato (abuso o negligencia), exposición prenatal a drogas o alcohol, hipoxia y hemorragia cerebral. Estos factores de riesgo para el desarrollo se han relacionado con puntuaciones más bajas obtenidas con distintas medidas de la aptitud cognitiva (p. ej., Baron et al., 2012; Raz, DeBastos, Newman y Batton, 2012; Rose, Feldman, Jankowski y Van Rossem, 2011). Algunos datos indican que los niños con factores de riesgo para el desarrollo muestran relativos puntos débiles en la memoria de trabajo y la velocidad de procesamiento (Rose et al., 2011).

En este estudio se esperaba que las medias de las puntuaciones compuestas obtenidas por los niños con factores de riesgo para el desarrollo fueran más bajas que las de los niños del grupo control pareado. Asimismo se predijo que las diferencias entre las medias del IMT, el IVP y el ICC obtenidas por los niños con factores de riesgo para el desarrollo y las medias correspondientes obtenidas por el grupo control ofrecerían tamaños del efecto más grandes que los del resto de puntuaciones compuestas. Se aplicó la WPPSI-IV a 42 niños de 2:6 a 7:6 años con un historial de factores de riesgo para el desarrollo. La tabla 6.19 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con factores de riesgo para el desarrollo, así como para el grupo control pareado.

Tabla 6.19. Resultados del grupo con factores de riesgo para el desarrollo y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ puntuación compuesta	Factores de riesgo para el desarrollo		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	7.7	4.2	10.3	2.1	42	2.64	3.73	<.01	0.80
Semejanzas	8.4	4.4	10.3	3.1	33	1.91	2.30	.03	0.50
Vocabulario	8.4	4.3	9.7	2.8	33	1.30	1.57	.13	0.36
Comprensión	8.0	4.6	10.0	3.2	33	1.97	2.09	.04	0.50
Dibujos	8.1	4.2	10.2	2.7	42	2.05	2.47	.02	0.58
Nombres	8.4	3.9	10.9	2.6	41	2.41	3.85	<.01	0.73
Cubos	8.0	4.3	10.1	2.5	41	2.12	2.87	<.01	0.60
Rompecabezas	7.8	3.7	10.5	2.2	42	2.62	4.16	<.01	0.86
Matrices	8.3	4.6	10.5	2.9	33	2.21	2.27	.03	0.57
Conceptos	8.2	4.9	10.1	2.3	33	1.91	2.10	.04	0.50
Reconocimiento	7.6	4.0	10.1	2.9	42	2.57	3.06	<.01	0.74
Localización	7.5	3.4	10.3	2.3	41	2.78	3.78	<.01	0.96
Búsqueda de animales	7.6	3.3	9.8	2.4	32	2.25	3.16	<.01	0.78
Cancelación	7.0	4.0	9.4	3.1	33	2.33	2.67	.01	0.65
Clave de figuras	7.4	4.3	9.2	2.4	33	1.88	2.34	.03	0.54
Cancelación aleatoria	7.0	4.0	9.6	3.3	33	2.58	2.94	<.01	0.70
Cancelación estructurada	7.2	4.1	9.3	2.8	33	2.12	2.48	.02	0.60
ICV	87.1	22.0	101.5	11.4	42	14.45	4.10	<.01	0.82
IVE	88.4	21.4	102.1	11.2	41	13.66	3.85	<.01	0.80
IRF	89.7	27.1	101.5	12.3	33	11.82	2.29	.03	0.56
IMT	85.8	19.6	101.3	12.5	41	15.54	3.72	<.01	0.95
IVP	85.1	19.3	97.8	12.7	32	12.63	3.14	<.01	0.77
CIT	85.9	22.4	101.5	9.5	40	15.65	4.13	<.01	0.91
IAV	90.2	21.4	103.0	12.1	41	12.80	3.53	<.01	0.74
INV	86.3	22.9	101.5	10.4	39	15.21	3.70	<.01	0.86
ICG	86.1	24.8	102.1	10.7	41	16.00	3.94	<.01	0.84
ICC	84.8	19.8	98.3	12.4	32	13.44	3.05	<.01	0.81

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

El grupo de niños con factores de riesgo para el desarrollo obtuvo puntuaciones compuestas significativamente inferiores a las del grupo control pareado. Las medias de los índices primarios del grupo con factores de riesgo para el desarrollo son de 85.1 (IVP) a 89.7 (IRF), y la media del CI total es de 85.9. Las medias de los índices secundarios oscilan entre 84.8 (ICC) y 90.2 (IAV). La media del INV es de 86.3, un valor muy similar al de la media del CI total. Los tamaños del efecto para las diferencias entre las medias de las puntuaciones compuestas son de moderados a grandes. Los tamaños del efecto más grandes se observan en el IMT, el CI total, el INV, el ICG y el ICV. Las diferencias entre las medias de las puntuaciones compuestas del grupo con factores de riesgo para el desarrollo y las del grupo control pareado no son tan grandes como las observadas en el grupo con discapacidad intelectual leve y su correspondiente grupo control pareado.

Todas las medias de las puntuaciones escalares de las pruebas obtenidas por el grupo con factores de riesgo para el desarrollo son significativamente más bajas que las del grupo control, con excepción de *Vocabulario*. Las pruebas de *Localización, Búsqueda de animales* y *Reconocimiento* ofrecen tres de los tamaños del efecto más grandes de las pruebas. Estos resultados son consistentes con estudios que habían indicado puntos débiles similares en la memoria de trabajo y la velocidad de procesamiento comparadas con otras áreas cognitivas (Rose et al., 2011). Se requieren más investigaciones sobre el rendimiento obtenido en la WPPSI-IV por los niños con factores de riesgo para el desarrollo, dadas las distintas etiologías. Además, puesto que el grupo con factores de riesgo para el desarrollo se consideró inicialmente como un grupo con factores de riesgo biológicos, es necesario confirmar con nuevos estudios si estos resultados se reproducen en muestras de niños con factores de riesgo socioambientales (p. ej., estatus económico bajo, poco acceso a los servicios).

Niños con problemas previos a la lectoescritura

Se supone que las destrezas previas a la lectoescritura, tales como el desarrollo del vocabulario, la sensibilidad fonológica y el conocimiento de las letras, son precursoras del desarrollo de la lectura. El nivel de desarrollo de estas destrezas antes de la enseñanza formal de la lectura aporta información sobre el posterior rendimiento lector y posibles trastornos de la lectura (Catts, Gillispie, Leonard, Kail y Miller, 2002; Lonigan, Burgess y Anthony, 2000). Muchos sujetos con problemas de la lectura pueden haber tenido un retraso del desarrollo del lenguaje oral (Mann, Cowin y Schoenheimer, 1989). Las destrezas del lenguaje oral son la base de las primeras etapas de la lectura. Por ejemplo, las dificultades en el vocabulario receptivo y expresivo y en la decodificación pueden dar lugar a una dificultad mucho mayor para la lectura de palabras y la comprensión de la lectura (Burns et al., 2002).

En este estudio se utilizó la prueba *Habilidades tempranas de la lectura* del WIAT-III para establecer los criterios de inclusión. Se predijo que los niños con problemas previos a la lectoescritura mostrarían un rendimiento significativamente más bajo que los niños del grupo control pareado en el ICG, CI total, INV e ICV de la WPPSI-IV. Además, dado que en la literatura se señalan las estrechas relaciones entre las destrezas lingüísticas y el inicio de la lectoescritura, se esperaba que el grupo con problemas previos a la lectoescritura obtuviera puntuaciones significativamente más bajas en el IAV que el grupo control.

Se aplicó la WPPSI-IV a 37 niños de 4:9 a 7:6 años con problemas previos a la lectoescritura. La tabla 6.20 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con problemas previos a la lectoescritura y para el grupo control pareado.

Tabla 6.20. Resultados del grupo con problemas previos a la lectoescritura y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ puntuación compuesta	Problemas previos a la lectoescritura		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	8.6	2.2	10.3	2.9	37	1.70	3.66	<.01	0.66
Semejanzas	8.0	2.4	10.3	3.4	37	2.32	3.98	<.01	0.79
Vocabulario	8.5	2.9	10.4	2.7	37	1.89	3.02	<.01	0.67
Comprensión	8.3	2.7	10.2	2.8	37	1.92	3.71	<.01	0.70
Dibujos	9.1	2.4	10.2	2.3	37	1.16	2.21	.03	0.49
Nombres	8.8	2.2	10.5	2.9	37	1.68	2.70	.01	0.65
Cubos	8.5	1.8	9.9	1.9	36	1.39	3.29	<.01	0.75
Rompecabezas	9.0	2.4	10.7	2.6	37	1.73	2.91	<.01	0.69
Matrices	8.6	2.4	10.5	2.4	37	1.84	2.93	<.01	0.77
Conceptos	8.5	3.0	10.1	2.5	37	1.59	2.37	.02	0.58
Reconocimiento	8.6	2.9	10.0	3.3	37	1.38	2.30	.03	0.44
Localización	9.0	2.3	9.8	2.2	37	0.86	1.68	.10	0.38
Búsqueda de animales	8.6	2.5	10.0	2.2	37	1.38	3.12	<.01	0.59
Cancelación	9.4	2.7	9.4	2.4	37	0.03	0.05	.96	0.01
Clave de figuras	8.6	2.2	9.8	2.2	37	1.19	2.52	.02	0.54
Cancelación aleatoria	9.4	2.8	9.5	2.5	37	0.05	0.09	.93	0.02
Cancelación estructurada	9.5	2.8	9.5	2.7	37	0.03	0.05	.96	0.01
ICV	89.9	9.7	101.5	15.3	37	11.51	4.40	<.01	0.90
IVE	92.8	9.9	102.4	9.9	36	9.56	4.02	<.01	0.97
IRF	91.8	13.4	101.6	10.8	37	9.86	3.23	<.01	0.81
IMT	93.0	11.8	99.8	11.4	37	6.84	2.84	<.01	0.59
IVP	94.5	12.8	98.5	11.2	37	4.00	1.71	.09	0.33
CIT	89.0	9.6	101.1	11.4	36	12.14	5.43	<.01	1.15
IAV	93.7	10.5	101.8	12.5	37	8.08	3.25	<.01	0.70
INV	89.9	11.2	101.1	9.3	36	11.19	5.11	<.01	1.09
ICG	89.4	9.2	101.3	11.7	36	11.92	5.09	<.01	1.13
ICC	92.3	12.0	98.8	9.8	37	6.54	2.97	<.01	0.60

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

El grupo con problemas previos a la lectoescritura mostró un rendimiento significativamente inferior al del grupo control en todas las puntuaciones compuestas, excepto en el IVP. Las medias de los índices primarios del grupo con problemas previos a la lectoescritura oscilan entre 89.9 (ICV) y 94.5 (IVP), y la media del CI total es de 89.0. Las medias de los índices secundarios oscilan entre 89.4 (ICG) y 93.7 (IAV). La media del INV es de 89.9, un valor muy similar al de la media del CI total. Los tamaños del efecto para las diferencias entre las puntuaciones compuestas son de moderados a grandes, excepto para el IVP, cuyo tamaño del efecto es pequeño. Los tamaños del efecto más grandes se observan en el CI total, el ICG, el INV, el IVE y el ICV.

Todas las puntuaciones escalares de las pruebas obtenidas por el grupo con problemas previos a la lectoescritura son significativamente más bajas que las del grupo control, con excepción de *Localización* y *Cancelación*. Los tamaños del efecto más grandes se observan en *Semejanzas, Matrices, Cubos, Comprensión, Rompecabezas, Vocabulario, Información* y *Nombres*.

Estos resultados son consistentes con estudios que habían observado, en los niños con problemas previos a la lectoescritura, puntos débiles relativos en el área del lenguaje, especialmente en el vocabulario expresivo y receptivo (Lonigan et al., 2000; Mann et al., 1989). El hecho de que el IVE presente el mayor tamaño del efecto de los índices primarios requiere cierto análisis. La ortografía es el aspecto visual de la lectura y es la base de las tareas de ordenación, lectura y deletreo de las letras que componen las palabras, tareas que se llevan a cabo mediante el conocimiento del

alfabeto, la memoria visual del patrón de letras y el reconocimiento rápido de palabras (Kendeou, Papadopoulos y Spanoudis, 2012). La atención visoespacial se ha asociado con el procesamiento ortográfico, un aspecto de la lectura (Waechter, Besner y Stolz, 2011); y el inicio de la lectura se ha vinculado recientemente con el desarrollo de la aptitud visoespacial en los niños pequeños (McBride-Chang et al., 2011). Dada la estrecha relación entre la aptitud previa a la lectoescritura y la aptitud visoespacial, no debe sorprender que el IVE ofrezca el mayor tamaño del efecto de todas las puntuaciones compuestas. Se requieren más investigaciones sobre el rendimiento obtenido en la WPPSI-IV por los niños con problemas previos a la lectoescritura, que han sido identificados mediante distintas metodologías (p. ej., derivación por parte del maestro).

Niños con trastorno por déficit de atención con hiperactividad

El trastorno por déficit de atención con hiperactividad (TDAH) es un trastorno del neurodesarrollo que se caracteriza por la disfunción de diversas regiones cerebrales y por problemas crónicos de atención (Hale et al., 2012). Aunque en algunos casos también pueden observarse conductas impulsivas e hiperactivas, los criterios del DSM-IV-TR no consideran los síntomas impulsivos e hiperactivos como requisito para diagnosticar TDAH. Dado que los problemas de atención están presentes en distintos trastornos propios de la infancia, los resultados de los tests pueden servir para respaldar un diagnóstico (Hale et al., 2012). Varios estudios indican que el TDAH se asocia con síntomas de falta de atención que coinciden, en gran medida, con los conceptos neuropsicológicos de *memoria de trabajo* y *función ejecutiva* (Hale et al., 2012; Martinussen, Hayden, Hogg-Johnson y Tannock, 2005).

Las puntuaciones tradicionales del CI total generalmente no han servido para discriminar a los niños o adultos con TDAH de la población no clínica; no obstante, algunos estudios indican que el funcionamiento intelectual global de los niños o adultos con TDAH refleja un leve deterioro (Hale et al., 2012). En cambio, las puntuaciones obtenidas por los niños con TDAH en tareas de comprensión verbal y de razonamiento perceptivo no son indicativas de ninguna alteración (Hale et al., 2012; Wechsler, 2002, 2003). Diversos estudios sugieren, además, que los niños con TDAH pueden presentar cierto deterioro de la memoria de trabajo y de la velocidad de procesamiento (Martinussen et al., 2005; Mayes, Calhoun, Chase, Mink y Stagg, 2009; Mayes, Calhoun, Mayes y Molitoris, 2012; Wechsler, 2003).

Tomando como referencia los resultados de las investigaciones previas, se esperaba que el grupo de niños con TDAH obtuviera puntuaciones medias significativamente inferiores a las del grupo control en el IMT, el IVP y el ICC de la WPPSI-IV. Se aplicó la WPPSI-IV a 53 niños de 3:1 a 7:7 años, que habían sido previamente identificados como niños con TDAH, de acuerdo con los criterios diagnósticos del DSM-IV-TR y los resultados significativos de las *Brown Attention-Deficit Disorder Scales for Children and Adolescents* (Brown, 2001) en su forma para los padres. El grupo de niños con TDAH incluía todos los subtipos de este trastorno (tipo con predominio del déficit de atención, tipo con predominio hiperactivo-impulsivo y tipo combinado). Los niños del grupo con TDAH debieron estar un mínimo de 24 horas antes de la aplicación de la WPPSI-IV sin tomar medicación psicoestimulante. Asimismo debían haber obtenido previamente una puntuación de la aptitud general (p. ej., el CI total) igual o superior a 80 en alguna medida estandarizada de la aptitud intelectual y de aplicación individual, o bien disponer de un informe en el que un psicólogo estimara una puntuación de la aptitud general igual o superior a 80. La tabla 6.21 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con TDAH y para el grupo control pareado.

Tabla 6.21. Resultados del grupo con TDAH y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ *puntuación compuesta*	Trastorno por déficit de atención con hiperactividad		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	*Media*	*Dt*	*n*	*Diferencia*	*t*	*p*	Diferencia típica[a]
Información	9.5	2.8	10.2	2.9	53	0.75	1.40	.17	0.26
Semejanzas	8.8	2.5	10.0	2.8	50	1.24	2.42	.02	0.47
Vocabulario	8.9	2.6	10.2	2.8	50	1.34	2.71	<.01	0.50
Comprensión	8.4	2.8	10.2	3.1	50	1.80	3.32	<.01	0.61
Dibujos	9.5	2.4	10.4	2.8	53	0.87	2.00	.05	0.33
Nombres	9.6	2.9	10.2	2.2	53	0.60	1.34	.18	0.23
Cubos	9.5	2.6	10.2	3.5	53	0.62	1.03	.31	0.20
Rompecabezas	10.2	3.1	10.5	2.6	53	0.25	0.42	.67	0.09
Matrices	9.2	2.4	10.1	3.0	50	0.84	1.41	.16	0.31
Conceptos	9.1	2.5	9.8	3.0	49	0.67	1.11	.27	0.24
Reconocimiento	9.2	2.6	10.4	2.8	53	1.23	2.27	.03	0.46
Localización	9.1	3.2	9.6	2.6	53	0.55	0.87	.39	0.19
Búsqueda de animales	8.7	3.2	9.8	2.5	50	1.14	2.18	.03	0.40
Cancelación	8.7	2.7	9.4	2.8	50	0.72	1.45	.15	0.26
Clave de figuras	8.5	3.0	10.2	2.9	50	1.72	3.02	<.01	0.58
Cancelación aleatoria	8.7	2.6	9.6	2.9	50	0.88	1.67	.10	0.32
Cancelación estructurada	9.0	2.8	9.4	3.0	50	0.42	0.84	.41	0.14
ICV	94.4	12.9	100.2	12.3	53	5.87	2.58	.01	0.47
IVE	99.6	13.8	102.0	15.7	53	2.40	0.80	.43	0.16
IRF	95.0	11.9	99.8	13.9	49	4.76	1.63	.11	0.37
IMT	94.9	14.1	100.0	13.6	53	5.09	1.81	.08	0.37
IVP	92.7	13.9	97.9	13.2	50	5.18	2.16	.04	0.38
CIT	93.4	12.1	100.5	13.4	53	7.06	2.99	<.01	0.55
IAV	97.3	12.8	101.5	12.2	53	4.19	1.97	.05	0.34
INV	93.6	12.4	100.3	14.2	52	6.69	2.57	.01	0.50
ICG	94.6	12.1	100.3	13.8	53	5.72	2.25	.03	0.44
ICC	92.9	13.4	98.9	11.7	50	6.00	2.72	<.01	0.48

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Aunque todas las medias de las puntuaciones compuestas obtenidas por el grupo con TDAH se sitúan en el rango promedio, se observan diferencias significativas entre las puntuaciones compuestas del grupo con TDAH y las del grupo control pareado, con excepción del IVE, el IRF y el IMT. La media más baja de todos los índices primarios para el grupo con TDAH se observa en el IVP. Las diferencias entre las medias del CI total y el INV muestran tamaños del efecto moderados. El resto de los índices primarios y secundarios muestran tamaños del efecto pequeños, excepto el IVE, cuyo tamaño del efecto es insignificante. Las diferencias entre las medias de las puntuaciones escalares de las pruebas obtenidas por el grupo con TDAH y el grupo control pareado son estadísticamente significativas para *Semejanzas, Vocabulario, Comprensión, Dibujos, Reconocimiento, Búsqueda de animales* y *Clave de figuras*. Los tamaños del efecto más grandes se observaron en *Comprensión, Clave de figuras, Vocabulario, Semejanzas, Reconocimiento* y *Búsqueda de animales*.

Los resultados de este estudio son consistentes con estudios previos que señalan que el rendimiento de los sujetos con TDAH en las tareas de memoria de trabajo y de velocidad de procesamiento es más bajo (Martinussen et al., 2005; Mayes et al., 2009; Mayes et al., 2012; Wechsler, 2003). El IVP, cuya media es la más baja de todos los índices primarios obtenidos por el grupo con TDAH, está estrechamente relacionado con la atención sostenida y la concentración. La

diferencia entre la media del ICV del grupo con el TDAH y la del grupo control es un resultado que no se esperaba y que puede indicar que algunos aspectos de la atención interfieren con la adquisición del conocimiento cristalizado. Hale et al. (2012) analizan la relación entre la función ejecutiva propia de los lóbulos frontales, que está alterada en el TDAH, y las tareas verbales que tienen una alta saturación en *g*. No obstante, en los estudios no se indica de una forma consistente que la aptitud verbal se encuentre afectada.

La diferencia significativa entre la media del CI total del grupo con TDAH y la del grupo control puede reflejar la mayor contribución de la memoria de trabajo y la velocidad de procesamiento en el CI total de la WPPSI-IV respecto al de la WPPSI-III. Esta observación no es consistente con la realizada por Hale et al. (2012); a saber, que el funcionamiento intelectual global de los niños o adultos con TDAH refleja un leve deterioro. Finalmente, dado que la muestra de este estudio con la WPPSI-IV incluía los tres tipos de TDAH (tipo con predominio del déficit de atención, tipo con predominio hiperactivo-impulsivo y tipo combinado), se requieren más investigaciones con muestras de niños en las que se diferencie cada tipo de TDAH y en las que los niños no hayan suspendido la toma de medicación psicoestimulante.

Niños con comportamiento perturbador

Los problemas conductuales pueden afectar al rendimiento académico de un niño, así como a su desarrollo social (Rudy y Levinson, 2008). Algunos de estos niños cumplen los criterios de la categoría de trastorno por comportamiento perturbador del DSM-IV-TR. Esta categoría incluye el trastorno disocial, el trastorno negativista desafiante y el trastorno de comportamiento perturbador no especificado. Estos niños muestran, en circunstancias normales, conductas inapropiadas que afectan negativamente a su funcionamiento escolar y social. Los problemas por comportamiento perturbador se han asociado con la aptitud verbal baja, los déficits en el vocabulario receptivo y la aptitud cognitiva general ligeramente baja (Barker et al., 2011; Dougherty et al., 2007; Isen, 2010; Séguin, Parent, Tremblay y Zelazo, 2009). Los trastornos por comportamiento perturbador parecen tener una compleja relación con funciones ejecutivas como la planificación, la organización y la memoria de trabajo (Drabick, Bubier, Chen, Price y Lanza, 2011; Raaijmakers et al., 2008). Una buena función ejecutiva no siempre indica menor sintomatología conductual; algunos estudios señalan justo lo contrario (Drabick et al., 2011).

Teniendo en cuenta los resultados de estudios previos, se esperaba que los niños del grupo con comportamiento perturbador obtuvieran puntuaciones medias más bajas en las pruebas de Comprensión verbal y en las pruebas de Memoria de trabajo que en el resto de las pruebas, así como que, en promedio, su media del CI total fuera baja. Se aplicó la WPPSI-IV a 26 niños de 4:7 a 7:2 años con problemas por comportamiento perturbador. Los niños del grupo con comportamiento perturbador habían sido diagnosticados con trastorno desafiante negativista o con trastorno disocial; o se había considerado que, en función de los criterios del centro escolar, presentaban un trastorno emocional en el que el comportamiento perturbador era el síntoma principal. Asimismo debían haber obtenido previamente una puntuación igual o superior a 80 en alguna medida estandarizada de la aptitud intelectual y de aplicación individual, o bien disponer de un informe en el que un psicólogo estimara una puntuación de la aptitud general igual o superior a 80. La tabla 6.22 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con comportamiento perturbador y para el grupo control pareado.

Tabla 6.22. Resultados del grupo con comportamiento perturbador y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ puntuación compuesta	Comportamiento perturbador		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	10.0	1.9	9.7	2.3	26	−0.31	−0.62	.54	−0.15
Semejanzas	8.8	2.4	9.8	2.5	25	0.92	1.37	.18	0.38
Vocabulario	8.1	3.0	9.0	2.9	26	0.92	1.15	.26	0.31
Comprensión	8.3	2.7	9.1	2.2	26	0.81	1.12	.27	0.33
Dibujos	9.2	2.3	9.5	2.3	26	0.31	0.43	.67	0.13
Nombres	8.0	2.8	9.8	2.9	26	1.81	1.88	.07	0.63
Cubos	9.5	2.5	10.0	2.9	26	0.54	0.63	.54	0.20
Rompecabezas	9.5	2.2	10.3	2.3	26	0.88	1.35	.19	0.39
Matrices	8.8	2.2	8.8	2.6	26	0.00	0.00	1.00	0.00
Conceptos	9.5	2.2	9.7	2.2	26	0.15	0.27	.79	0.07
Reconocimiento	8.7	3.0	9.0	2.8	26	0.27	0.36	.72	0.09
Localización	8.6	2.3	9.4	2.4	26	0.85	1.37	.18	0.36
Búsqueda de animales	7.5	2.9	9.6	3.3	26	2.08	2.28	.03	0.67
Cancelación	9.8	3.4	9.5	2.9	26	−0.31	−0.36	.72	−0.10
Clave de figuras	9.2	3.8	9.2	2.9	26	−0.08	−0.07	.94	−0.02
Cancelación aleatoria	10.1	3.4	9.7	2.9	26	−0.42	−0.50	.62	−0.13
Cancelación estructurada	9.8	3.4	9.6	3.3	26	−0.15	−0.17	.87	−0.04
ICV	96.3	10.7	98.4	11.1	25	2.08	0.72	.48	0.19
IVE	97.2	10.5	101.2	12.5	26	4.00	1.13	.27	0.35
IRF	95.0	10.3	95.4	12.3	26	0.35	0.11	.91	0.03
IMT	92.0	13.0	95.2	12.6	26	3.15	0.94	.36	0.25
IVP	92.6	15.6	97.7	14.7	26	5.12	1.16	.26	0.34
CIT	91.8	10.6	96.7	12.1	25	4.88	1.62	.12	0.43
IAV	92.2	12.6	98.2	12.8	26	5.96	1.37	.18	0.47
INV	91.7	11.9	95.7	12.1	26	4.04	1.31	.20	0.34
ICG	95.4	9.3	97.6	11.6	25	2.24	0.80	.43	0.21
ICC	90.7	13.4	95.9	13.9	26	5.19	1.45	.16	0.38

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Todas las medias de las puntuaciones compuestas obtenidas por el grupo con comportamiento perturbador se sitúan en el rango promedio, y la más baja (90.7) corresponde al ICC. La media del CI total de este grupo es de 91.8. Las diferencias entre las medias de todas las puntuaciones compuestas obtenidas por el grupo con comportamiento perturbador y las obtenidas por el grupo control ofrecen tamaños del efecto pequeños, con excepción del ICV y del IRF con tamaños del efecto insignificantes. Y los tamaños del efecto más grandes se observan en el IAV, el CI total y el ICC. En cuanto a las pruebas, solo la diferencia entre las medias de la puntuación escalar de *Búsqueda de animales* es significativa. Los tamaños del efecto observados en *Nombres* y *Búsqueda de animales* son moderados. Las diferencias entre las medias de las puntuaciones escalares de *Semejanzas, Vocabulario, Comprensión, Cubos, Rompecabezas* y *Localización* muestran tamaños del efecto pequeños.

Los resultados de este estudio coinciden, en parte, con estudios previos que señalan que los niños con comportamiento perturbador obtienen puntuaciones relativamente más bajas en algunas tareas verbales (IAV y *Nombres*, pero no ICV) y de memoria de trabajo (*Localización*, pero no *Reconocimiento*), así como que su aptitud cognitiva general es ligeramente inferior a la de la población general (Barker et al., 2011; Dougherty et al., 2007; Isen, 2010; Séguin et al., 2009). Aunque el tamaño del efecto moderado obtenido en *Búsqueda de animales* es un valor que no se esperaba, cabe tener en cuenta que, en los casos que presentan un deterioro de la memoria de

trabajo o de otra función ejecutiva, se observa con cierta frecuencia un rendimiento más bajo en las tareas de velocidad de procesamiento. Esta observación requiere más investigaciones que reproduzcan las características de este estudio. Además, dado que la muestra de este estudio estaba compuesta por niños con distintos diagnósticos, es necesario realizar más investigaciones para determinar el rendimiento en la WPPSI-IV por parte de grupos más específicos de niños con comportamiento perturbador.

Niños con trastorno del lenguaje expresivo

De acuerdo con el DSM-IV-TR, los niños diagnosticados con trastorno del lenguaje expresivo obtienen puntuaciones más bajas en las medidas del desarrollo del lenguaje expresivo que en las medidas del desarrollo de la inteligencia no verbal y del lenguaje receptivo. Dado que los niños que formaron parte de este estudio habían sido diagnosticados según los criterios del DSM-IV-TR y los resultados de estudios previos (Raiford et al., 2005; Wechsler, 2002, 2003), se esperaba que sus puntuaciones medias en el ICV, el CI total y el ICG fueran significativamente más bajas que las del grupo control pareado. Asimismo se predijo que obtendrían puntuaciones escalares medias significativamente más bajas en las pruebas que requieren respuestas verbales expresivas (*Información, Semejanzas, Vocabulario, Comprensión* y *Nombres*).

Se aplicó la WPPSI-IV a 25 niños de 4:0 a 7:6 años, que habían sido diagnosticados con trastorno del lenguaje expresivo, de acuerdo con los criterios del DSM-IV-TR. La tabla 6.23 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con trastorno del lenguaje expresivo y para el grupo control pareado.

Tabla 6.23. Resultados del grupo con trastorno del lenguaje expresivo y del grupo control

Prueba/puntuación *de procesamiento*/ puntuación compuesta	Trastorno del lenguaje expresivo		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	7.7	2.3	10.3	2.6	23	2.61	4.22	<.01	1.06
Semejanzas	7.2	2.8	10.6	2.5	25	3.48	4.04	<.01	1.31
Vocabulario	7.6	2.5	10.5	2.5	25	2.92	4.39	<.01	1.17
Comprensión	7.3	2.0	10.1	3.0	25	2.84	5.38	<.01	1.11
Dibujos	9.2	2.4	10.2	3.1	25	1.00	1.34	.19	0.36
Nombres	8.2	2.2	10.6	2.3	25	2.36	3.48	<.01	1.05
Cubos	9.5	2.4	10.0	3.0	25	0.56	0.69	.50	0.21
Rompecabezas	9.8	2.1	10.3	2.7	25	0.56	0.81	.42	0.23
Matrices	9.9	3.5	10.6	2.6	25	0.76	0.85	.40	0.25
Conceptos	8.7	3.0	10.8	2.4	25	2.08	3.01	<.01	0.77
Reconocimiento	8.4	3.1	9.8	3.2	25	1.36	2.18	.04	0.43
Localización	8.4	2.4	9.8	2.5	25	1.48	1.92	.07	0.60
Búsqueda de animales	8.6	2.7	10.1	2.7	25	1.52	1.89	.07	0.56
Cancelación	9.2	2.7	9.8	2.3	25	0.68	0.98	.34	0.27
Clave de figuras	8.6	4.0	10.0	2.5	25	1.40	1.49	.15	0.42
Cancelación aleatoria	8.9	2.7	9.5	1.9	25	0.60	0.89	.38	0.26
Cancelación estructurada	9.4	2.7	10.3	2.5	25	0.84	1.07	.30	0.32
ICV	86.1	12.1	102.6	13.2	23	16.48	4.33	<.01	1.30
IVE	98.0	11.0	101.2	13.8	25	3.20	0.89	.38	0.26
IRF	95.7	15.2	104.0	11.3	25	8.28	2.27	.03	0.62
IMT	90.7	13.9	99.0	13.3	25	8.24	2.62	.02	0.61
IVP	93.8	12.0	100.1	12.5	25	6.24	1.90	.07	0.51
CIT	89.7	11.3	102.3	14.1	23	12.70	4.18	<.01	0.99
IAV	92.4	10.0	102.0	12.0	25	9.56	3.63	<.01	0.87
INV	93.1	12.4	102.0	13.9	25	8.88	2.72	.01	0.67
ICG	90.5	12.4	103.0	12.4	23	12.43	3.82	<.01	1.00
ICC	90.5	12.1	99.4	13.2	25	8.88	2.79	.01	0.70

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

En cuanto a las puntuaciones compuestas, las diferencias más altas entre las medias de ambos grupos se observan en el ICV, el ICG y el CI total, que también muestran los tamaños del efecto más grandes. El tamaño del efecto del IAV también es grande. Los tamaños del efecto son moderados para el IRF, el IMT, el IVP, el INV y el ICC. La media del INV del grupo con trastorno del lenguaje expresivo (93.1) es ligeramente más alta que la media del CI total (89.7). El IVE muestra un tamaño del efecto pequeño.

Respecto a las pruebas, las puntuaciones escalares de todas las pruebas que requieren respuestas expresivas, así como las de *Conceptos* y *Reconocimiento*, son significativamente más bajas en el grupo con trastorno del lenguaje expresivo que en el grupo control. De las pruebas que muestran diferencias significativas entre las medias, los mayores tamaños del efecto se observan en *Semejanzas, Vocabulario, Comprensión, Información, Nombres* y *Conceptos*. Las pruebas *Localización* y *Búsqueda de animales* muestran tamaños del efecto moderados, y el resto de las pruebas ofrecen tamaños del efecto pequeños. La diferencia entre las medias obtenidas por ambos grupos en *Dibujos* muestra el tamaño del efecto más bajo de todas las pruebas de Comprensión verbal, lo que refleja que en esta prueba no interviene el lenguaje expresivo.

Estos resultados son consistentes con estudios previos basados en niños con trastorno del lenguaje expresivo, en los que se observa que su rendimiento es bajo en las tareas de comprensión verbal y de razonamiento verbal que requieren expresión, y que sus puntuaciones compuestas más bajas

son las de los índices formados, en parte o completamente, por dichas tareas (Raiford et al., 2005; Wechsler, 2002, 2003). Asimismo, estos resultados concuerdan con estudios que señalan que el rendimiento de los niños con trastornos del lenguaje es bajo en las tareas de memoria de trabajo visual y de velocidad de procesamiento (Bavin, Wilson, Maruff y Sleeman, 2005; Cardy, Tannock, Johnson y Johnson, 2010; Marton, 2008).

La media del INV correspondiente al grupo con trastorno del lenguaje expresivo es ligeramente más alta que la del CI total, aun así es significativamente más baja que la media del INV del grupo control. Aunque el INV esté compuesto exclusivamente de tareas que no requieren respuestas expresivas, depende, en parte, de aptitudes que también son sensibles al trastorno del lenguaje expresivo, tales como el razonamiento fluido, la memoria de trabajo y la velocidad de procesamiento. Por lo tanto, aun dependiendo menos de los principales puntos débiles asociados con el trastorno del lenguaje expresivo, el INV se ve afectado por puntos débiles cognitivos secundarios pero relacionados. La media más alta de las puntuaciones compuestas obtenidas por el grupo con trastorno del lenguaje expresivo es el IVE, que parece estar menos influenciado por dichos puntos débiles cognitivos secundarios; y, que, por tanto, es la medida de un dominio cognitivo con menos afectación.

En conjunto, estos resultados demuestran que la WPPSI-IV es una herramienta útil para la detección del trastorno del lenguaje expresivo. Se requieren más investigaciones con muestras más amplias de niños con trastorno del lenguaje expresivo para determinar la utilidad clínica del INV en esta población, pero la demostración inicial parece respaldar la utilidad del INV para evaluar este trastorno.

Niños con trastorno mixto del lenguaje receptivo-expresivo

Los niños diagnosticados con trastorno mixto del lenguaje receptivo-expresivo obtienen puntuaciones considerablemente más bajas en las medidas del lenguaje receptivo y expresivo, en comparación con las medidas no verbales de inteligencia (DSM-IV-TR). Estos niños suelen mostrar un rendimiento más bajo en las tareas de comprensión verbal, razonamiento perceptivo, memoria de trabajo y velocidad de procesamiento; puntuaciones más bajas en el LG (actualmente denominado *IAV*) y en el CI total que los niños sin trastornos del lenguaje; así como tamaños del efecto relativamente mayores para las diferencias entre las medias del ICV, el IVP, el CI total y el LG (p. ej., Archibald y Gathercole, 2006a; Bavin et al., 2005; Cardy et al., 2010; Marton, 2008; Wechsler, 2002, 2003).

Tomando como referencia las investigaciones previas, se esperaba que el grupo con trastorno mixto del lenguaje receptivo-expresivo obtuviera puntuaciones medias significativamente más bajas que el grupo control pareado en la mayoría de las puntuaciones compuestas, sino en todas, así como tamaños de efecto más grandes para las diferencias entre las medias del ICV, el IMT, el IVP, el CI total, el IAV, el ICG y el ICC que para las puntuaciones del IVE. En cuanto a las pruebas, se predijo que las puntuaciones escalares de todas las pruebas de Comprensión verbal y de Velocidad de procesamiento serían significativamente más bajas en el grupo con trastorno mixto del lenguaje receptivo-expresivo que en el grupo control pareado. Se aplicó la WPPSI-IV a 42 niños de 4:0 a 7:6 años, que habían sido diagnosticados con trastorno mixto del lenguaje receptivo-expresivo, de acuerdo con los criterios del DSM-IV-TR. La tabla 6.24 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con trastorno mixto del lenguaje receptivo-expresivo y para el grupo control pareado.

Tabla 6.24. Resultados del grupo con trastorno mixto del lenguaje receptivo-expresivo y del grupo control

Prueba/puntuación de procesamiento/ puntuación compuesta	Trastorno mixto del lenguaje receptivo-expresivo		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	5.3	2.8	9.9	2.7	39	4.56	7.25	<.01	1.66
Semejanzas	6.6	2.7	10.1	2.5	42	3.52	7.27	<.01	1.35
Vocabulario	6.0	2.6	9.7	2.7	39	3.74	6.28	<.01	1.41
Comprensión	5.8	2.4	9.7	2.9	39	3.85	7.45	<.01	1.45
Dibujos	6.8	2.9	10.3	2.7	41	3.51	5.66	<.01	1.25
Nombres	6.5	2.9	9.7	3.0	40	3.13	4.75	<.01	1.06
Cubos	7.9	3.1	9.6	2.5	42	1.69	3.64	<.01	0.60
Rompecabezas	8.9	3.6	9.8	2.8	42	0.88	1.21	.23	0.27
Matrices	7.5	3.0	9.8	3.1	42	2.33	3.66	<.01	0.76
Conceptos	7.7	2.8	9.8	2.7	42	2.10	3.88	<.01	0.76
Reconocimiento	8.0	2.3	9.8	2.6	42	1.71	3.42	<.01	0.70
Localización	8.0	3.3	10.4	3.1	41	2.39	3.16	<.01	0.75
Búsqueda de animales	7.3	2.9	10.5	2.9	42	3.21	5.82	<.01	1.11
Cancelación	7.9	4.2	10.1	2.9	42	2.21	2.73	<.01	0.61
Clave de figuras	8.2	3.5	10.1	3.4	41	1.90	2.79	<.01	0.55
Cancelación aleatoria	8.0	4.2	10.2	2.7	42	2.21	2.77	<.01	0.63
Cancelación estructurada	7.9	4.2	10.1	3.0	42	2.24	2.70	<.01	0.61
ICV	78.3	12.3	99.3	13.1	39	21.03	7.95	<.01	1.66
IVE	91.4	15.8	98.6	13.6	42	7.19	2.50	.02	0.49
IRF	85.9	13.5	98.6	13.9	42	12.74	4.70	<.01	0.93
IMT	89.0	13.9	100.8	14.7	41	11.78	3.82	<.01	0.82
IVP	86.8	17.7	102.1	14.4	42	15.31	4.79	<.01	0.95
CIT	79.5	12.3	98.9	13.3	39	19.38	8.33	<.01	1.51
IAV	81.8	14.0	99.4	13.6	39	17.64	5.45	<.01	1.28
INV	84.1	14.5	99.3	13.8	42	15.19	6.28	<.01	1.07
ICG	78.8	12.7	98.3	13.0	39	19.49	8.29	<.01	1.52
ICC	86.3	15.9	101.9	14.5	41	15.59	4.64	<.01	1.02

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Tal como se esperaba, el grupo con trastorno mixto del lenguaje receptivo-expresivo obtuvo todas las puntuaciones compuestas significativamente más bajas que el grupo control, así como tamaños del efecto grandes para todas las diferencias entre las medias, excepto para las del IVE. Los tamaños del efecto más grandes se observaron en el ICV, el ICG, el CI total y el IAV. La media del INV del grupo con trastorno mixto del lenguaje receptivo-expresivo (84.1) es ligeramente más alta que la media del CI total (79.5). El tamaño del efecto para el IVE es pequeño, en contraste con el tamaño del efecto para el IRF, que es grande. En cuanto a las pruebas, las puntuaciones escalares de todas las pruebas son significativamente más bajas en el grupo con trastorno mixto del lenguaje receptivo-expresivo que en el grupo control pareado, con excepción de *Rompecabezas*, aunque se aprecia que la diferencia entre las medias de esta prueba tiene un efecto pequeño. De todas las pruebas, las que muestran los tamaños del efecto más grandes son *Información, Comprensión, Vocabulario, Semejanzas, Dibujos, Búsqueda de animales* y *Nombres*.

Estos resultados son consistentes con los estudios que sugieren que los niños con trastornos del lenguaje tienden a presentar déficits globales en el funcionamiento cognitivo, un rendimiento más alto en las tareas visoespaciales respecto a otras tareas, un rendimiento bajo en memoria de trabajo y una velocidad de procesamiento más baja respecto a los niños sin problemas del lenguaje (Archibald y Gathercole, 2006a; Bavin et al., 2005; Cardy et al., 2010; Marton, 2008; Wechsler,

2002, 2003). Además, estos resultados aportan datos preliminares sobre la utilidad clínica del IVE y del IRF por separado para detectar el trastorno mixto del lenguaje receptivo-expresivo.

Al igual que en el grupo con trastorno del lenguaje expresivo, la media del INV correspondiente al grupo con trastorno del lenguaje receptivo-expresivo es ligeramente más alta que la del CI total, aun así es significativamente más baja que la media del INV del grupo control. Aunque el INV esté compuesto exclusivamente de tareas que no requieren respuestas expresivas, depende, en parte, de aptitudes que también son sensibles al trastorno mixto del lenguaje receptivo-expresivo, tales como el razonamiento fluido, la memoria de trabajo y la velocidad de procesamiento. Por lo tanto, aun dependiendo menos de los principales puntos débiles asociados con el trastorno mixto del lenguaje receptivo-expresivo, el INV se ve afectado por puntos débiles cognitivos secundarios pero relacionados. Se requieren más investigaciones con muestras más amplias de niños con trastorno mixto del lenguaje receptivo-expresivo para determinar la utilidad clínica del INV en esta población, pero la demostración inicial parece respaldar la utilidad del INV para evaluar este trastorno. Además, al igual que en el grupo con trastorno del lenguaje expresivo, la media más alta de las puntuaciones compuestas obtenidas por el grupo con trastorno mixto del lenguaje receptivo-expresivo es el IVE, quizá porque está menos influenciado por dichos puntos débiles cognitivos secundarios.

Niños con trastorno autista

Los niños con trastorno autista se caracterizan por presentar déficits en la comunicación verbal y no verbal y en la interacción social, así como movimientos estereotipados, acciones repetitivas y resistencia al cambio, de acuerdo con el DSM-IV-TR. Los cambios propuestos para los criterios diagnósticos del trastorno autista lo reconceptualizan como un trastorno dentro de un espectro. Los criterios propuestos requieren especificar la gravedad del diagnóstico (Klinger, O'Kelley, Mussey, Goldstein y DeVries, 2012). Diversas investigaciones previas sugieren que el funcionamiento intelectual general de los niños con trastorno autista es inferior al de los niños del grupo control. Típicamente, el rendimiento en las tareas verbales (p. ej. pruebas de Comprensión verbal) es más bajo para la mayoría de los niños con trastorno autista (Klinger et al., 2012; Wechsler, 2002, 2003). Algunos estudios señalan puntos fuertes relativos para estos niños en las tareas visoespaciales (Mayes y Calhoun, 2008; Wechsler, 2002, 2003).

Tomando en consideración investigaciones anteriores, se esperaba que todas las medias de las puntuaciones compuestas obtenidas por los niños con trastorno autista fueran significativamente más bajas que las del grupo control. También se predijo que la diferencia entre las medias del ICV ofrecería tamaños del efecto más grandes que las diferencias del resto de índices primarios.

Se aplicó la WPPSI-IV a 38 niños de 2:10 a 7:6 años, que habían sido diagnosticados con trastorno autista, de acuerdo con los criterios del DSM-IV-TR. Se utilizaron los criterios del DSM-IV-TR porque los criterios revisados no se habían publicado cuando se definió la muestra de este estudio. Se excluyó del estudio a los niños con trastorno autista que habían obtenido puntuaciones en pruebas de aptitud cognitiva general de 2.67 desviaciones típicas o más por debajo de la media (p. ej., CI total < 60). La sobrerrepresentación de niños respecto a las niñas pretende reflejar el efecto combinado de la mayor tasa de prevalencia de varones respecto a mujeres en el trastorno autista, así como la mayor prevalencia de discapacidad intelectual en las mujeres con trastorno autista (DSM-IV-TR). La tabla 6.25 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con trastorno autista y para el grupo control pareado.

Tabla 6.25. Resultados del grupo con trastorno autista y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ puntuación compuesta	Trastorno autista		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	4.9	2.6	10.5	2.4	38	5.66	10.71	<.01	2.26
Semejanzas	5.5	2.9	10.7	2.4	34	5.26	9.76	<.01	1.98
Vocabulario	5.2	2.5	10.8	3.1	34	5.59	9.60	<.01	1.99
Comprensión	4.5	2.6	10.7	2.7	33	6.12	9.69	<.01	2.31
Dibujos	5.8	2.3	10.1	3.2	38	4.34	6.78	<.01	1.56
Nombres	6.5	3.3	10.4	2.8	38	3.95	6.53	<.01	1.29
Cubos	8.0	3.3	10.3	2.9	37	2.30	3.11	<.01	0.74
Rompecabezas	7.6	3.5	10.2	2.4	38	2.61	3.44	<.01	0.87
Matrices	7.7	3.2	9.9	2.9	34	2.18	3.14	<.01	0.71
Conceptos	6.7	3.0	10.3	3.5	34	3.59	4.12	<.01	1.10
Reconocimiento	7.3	3.5	9.6	2.6	37	2.27	2.76	<.01	0.74
Localización	7.3	3.4	10.4	2.9	38	3.13	4.32	<.01	0.99
Búsqueda de animales	5.7	3.2	10.1	2.5	34	4.35	6.37	<.01	1.51
Cancelación	4.8	3.4	9.8	2.5	34	4.94	6.55	<.01	1.66
Clave de figuras	6.7	3.3	10.4	3.0	33	3.70	5.04	<.01	1.17
Cancelación aleatoria	5.1	3.6	9.7	2.5	34	4.59	6.29	<.01	1.48
Cancelación estructurada	4.9	3.1	10.0	2.3	34	5.06	6.86	<.01	1.85
ICV	75.2	10.9	102.7	11.3	38	27.45	12.46	<.01	2.47
IVE	87.6	17.0	101.7	13.2	37	14.11	3.71	<.01	0.93
IRF	83.9	15.3	100.6	15.4	34	16.71	4.33	<.01	1.09
IMT	84.1	16.7	99.6	13.6	37	15.57	3.94	<.01	1.02
IVP	73.7	15.9	99.6	11.0	34	25.91	7.83	<.01	1.90
CIT	77.6	12.4	100.4	11.5	36	22.83	8.07	<.01	1.91
IAV	78.9	13.4	101.6	13.9	38	22.63	8.08	<.01	1.66
INV	80.9	14.6	99.0	13.2	36	18.08	5.10	<.01	1.30
ICG	78.5	13.1	101.5	12.0	37	22.97	8.51	<.01	1.83
ICC	75.9	16.3	99.2	8.1	33	23.30	6.74	<.01	1.81

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Para el grupo con trastorno autista, las medias de los índices primarios de la WPPSI-IV oscilan entre 73.7 (IVP) y 87.6 (IVE), y la media del CI total es de 77.6. Todas estas medias son significativamente más bajas que las medias correspondientes del grupo control. Y los tamaños del efecto para las diferencias entre las medias de las puntuaciones compuestas fueron grandes. Tal como se esperaba, en el grupo con trastorno autista, el IVE es más alto que el ICV. Los índices secundarios de este grupo oscilan entre 75.9 (ICC) y 80.9 (INV), y la media del INV es ligeramente más alta que la media del CI total. El tamaño del efecto del INV, aun siendo grande, es ligeramente más pequeño que el del CI total. En cuanto a las pruebas, todas las medias de las puntuaciones escalares son significativamente más bajas en el grupo con trastorno autista que en el grupo control pareado. Los mayores tamaños del efecto se observan en *Comprensión, Información, Vocabulario* y *Semejanzas*.

Estos resultados coinciden con los de investigaciones anteriores que indican déficits cognitivos globales, un rendimiento relativamente más bajo en las tareas verbales y un rendimiento relativamente más alto en las tareas visoespaciales (Klinger et al., 2012; Mayes y Calhoun, 2008; Wechsler, 2002, 2003). Estos resultados demuestran la utilidad de la WPPSI-IV como medida de aptitud general que debe formar parte de una evaluación completa de los niños con trastorno autista. Se requieren más investigaciones para determinar si estos resultados son similares a los obtenidos por niños con trastorno autista y niveles inferiores de aptitud intelectual general.

Niños con trastorno de Asperger

El diagnóstico diferencial entre el trastorno autista y el trastorno de Asperger depende principalmente de los niveles de aptitud lingüística general y de interacción social, de acuerdo con los criterios del DSM-IV-TR. Los niños diagnosticados con trastorno de Asperger muestran déficits menos graves en la aptitud verbal y menos retraso del desarrollo cognitivo, en cuanto al aprendizaje y la conducta adaptativa, que los niños con trastorno autista. Los cambios propuestos en los criterios diagnósticos clasifican el trastorno de Asperger dentro del trastorno del espectro autista. Los criterios propuestos requieren especificar la gravedad del diagnóstico; por ello, algunos autores utilizan indistintamente los términos *trastorno de Asperger* y *trastorno del espectro autista de alto funcionamiento* (p. ej., Klinger et al., 2012). En el momento de redactar este manual se seguían analizando estos criterios a fin de examinar su sensibilidad y especificidad en relación con los criterios del DSM-IV-TR (p. ej., McPartland, Reichow y Volkmar, 2012; Swedo et al., 2012).

Diversos estudios previos señalan que los sujetos con trastorno de Asperger muestran puntos débiles en la velocidad de procesamiento y conservan la aptitud de comprensión verbal, aunque obtienen puntuaciones más bajas en la prueba *Comprensión* que en el resto de las pruebas de Comprensión verbal (Mayes y Calhoun, 2008; Mayes et al., 2012; Wechsler, 2003, 2008). Otros estudios también indican que, en estos sujetos, el ICG suele ser más alto que el CI total (Raiford et al., 2005; Wechsler, 2008). En este estudio de la WPPSI-IV se esperaba obtener resultados similares; además, se predijo que la media del ICC sería ligeramente más baja que la media del ICG.

Se aplicó la WPPSI-IV a 38 niños de 3:10 a 7:6 años, que habían sido diagnosticados con trastorno de Asperger, de acuerdo con los criterios del DSM-IV-TR. Se utilizaron los criterios del DSM-IV-TR porque los criterios revisados no se habían publicado cuando se definió la muestra de este estudio. Se excluyó del estudio a los niños con trastorno de Asperger que habían obtenido puntuaciones en pruebas de aptitud cognitiva general de 2 desviaciones típicas o más por debajo de la media (p. ej., CI total < 70). La tabla 6.26 presenta las medias de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas de la WPPSI-IV, para el grupo de niños con trastorno de Asperger y para el grupo control pareado.

Tabla 6.26. Resultados del grupo con trastorno de Asperger y del grupo control (datos estadounidenses)

Prueba/puntuación de procesamiento/ puntuación compuesta	Trastorno de Asperger		Grupo control		Diferencia de medias de los grupos				
	Media	*Dt*	Media	*Dt*	*n*	Diferencia	*t*	*p*	Diferencia típica[a]
Información	10.8	2.8	10.4	2.4	38	−0.34	−0.60	.55	−0.13
Semejanzas	10.6	3.1	10.6	2.1	37	0.05	0.09	.93	0.02
Vocabulario	8.3	3.3	10.9	2.3	37	2.65	3.81	<.01	0.93
Comprensión	7.4	3.0	10.6	2.8	37	3.24	4.43	<.01	1.12
Dibujos	9.7	3.7	10.4	2.0	37	0.73	1.18	.25	0.25
Nombres	11.0	3.7	10.7	2.5	38	−0.26	−0.36	.72	−0.08
Cubos	10.9	3.1	10.4	3.4	37	−0.49	−0.66	.51	−0.15
Rompecabezas	10.0	3.0	10.0	2.5	38	0.00	0.00	1.00	0.00
Matrices	10.4	2.8	10.1	2.8	37	−0.30	−0.40	.69	−0.11
Conceptos	9.4	2.5	10.6	2.3	37	1.14	1.93	.06	0.47
Reconocimiento	9.6	3.1	9.7	2.6	38	0.11	0.18	.86	0.04
Localización	9.5	2.8	9.6	2.4	38	0.13	0.22	.83	0.05
Búsqueda de animales	8.6	3.2	9.6	2.6	37	0.92	1.76	.09	0.32
Cancelación	8.8	3.6	9.3	2.5	37	0.51	0.82	.42	0.16
Clave de figuras	9.1	2.8	10.0	3.1	37	0.92	1.43	.16	0.31
Cancelación aleatoria	8.8	3.4	9.4	2.8	37	0.59	0.91	.37	0.19
Cancelación estructurada	8.8	3.7	9.1	2.6	37	0.32	0.49	.62	0.10
ICV	103.1	15.3	102.2	11.0	38	−0.89	−0.32	.75	−0.07
IVE	103.1	15.9	101.7	15.1	37	−1.43	−0.39	.70	−0.09
IRF	99.4	14.1	101.8	12.2	37	2.41	0.73	.47	0.18
IMT	97.3	15.0	97.9	12.5	38	0.58	0.19	.85	0.04
IVP	93.0	17.3	96.8	12.7	37	3.78	1.34	.19	0.25
CIT	101.3	15.0	101.1	12.4	37	−0.16	−0.06	.95	−0.01
IAV	101.5	18.5	103.2	10.8	37	1.70	0.51	.61	0.11
INV	98.5	14.3	100.6	12.1	37	2.05	0.74	.47	0.15
ICG	104.1	14.7	102.3	13.0	37	−1.78	−0.60	.55	−0.13
ICC	94.2	16.0	96.8	11.5	37	2.68	0.93	.36	0.19

[a] La diferencia típica es la diferencia entre las medias de las dos aplicaciones dividida por la raíz cuadrada de la varianza común, calculada mediante la fórmula de Cohen (1996, fórmula 10.4).

Para el grupo con trastorno de Asperger, las medias de los índices primarios de la WPPSI-IV oscilan entre 93.0 (IVP) y 103.1 (ICV e IVE), y la media del CI total es de 101.3. Las diferencias entre las medias de las puntuaciones compuestas del grupo con trastorno de Asperger y del grupo control no son significativas; no obstante, el tamaño del efecto para la diferencia entre las medias del IVP es pequeño. Los índices secundarios de la WMS-IV para el grupo de sujetos con trastorno de Asperger oscilan entre 94.2 (ICC) y 104.1 (ICG). Tal como se esperaba, en el grupo con trastorno de Asperger, el ICC es más bajo que el ICG. En cuanto a las pruebas, las medias de las puntuaciones escalares de *Vocabulario* y *Comprensión* son significativamente más bajas en el grupo con trastorno de Asperger que en el grupo control. Los mayores tamaños del efecto se observan en *Comprensión* y *Vocabulario*. Las diferencias entre las medias de *Dibujos, Conceptos, Búsqueda de animales* y *Clave de figuras* no son estadísticamente significativas, pero ofrecen tamaños del efecto pequeños.

Estos datos son consistentes con los criterios del DSM-IV-TR y con investigaciones previas que señalan déficits menos graves en la aptitud verbal de los niños con trastorno de Asperger que en la de los niños con trastorno autista. El resultado obtenido por el grupo con trastorno de Asperger en la prueba *Vocabulario* es excepcional y requiere una investigación más detallada. Para explicar este resultado, pueden proponerse algunas hipótesis. En primer lugar, es posible que el rango de edad de la WPPSI-IV, en relación con muestras de niños más mayores, adolescentes o adultos (p. ej.,

Wechsler, 2003, 2008), se asocie con una mayor incertidumbre diagnóstica respecto al desarrollo del lenguaje, lo que dificultaría la elección entre el trastorno autista y el trastorno de Asperger. En segundo lugar, tal vez los nuevos criterios propuestos para el trastorno del espectro autista dio lugar a que algunos profesionales combinaran los diagnósticos. No obstante, las puntuaciones medias obtenidas por el grupo con trastorno de Asperger en la prueba *Vocabulario* y en otras pruebas verbales son más altas que las del grupo con trastorno autista, lo que indica que los niños del grupo con trastorno de Asperger conservan relativamente la aptitud de comprensión verbal. Se requieren más investigaciones para determinar si este patrón de resultados se reproduce con muestras de sujetos con trastorno de Asperger con niveles inferiores de aptitud intelectual general.

Resumen del rendimiento de los grupos especiales en la WPPSI-IV

Los resultados obtenidos por los grupos especiales respaldan la validez y utilidad clínica de la WPPSI-IV, y generalmente son consistentes con las expectativas basadas en estudios anteriores y en fundamentos teóricos. Se espera que las próximas investigaciones basadas en la WPPSI-IV y en ambientes y poblaciones clínicas diferentes aporten más datos sobre la utilidad de la escala como parte de una exhaustiva evaluación clínica del desarrollo llevada a cabo con fines diagnósticos y para definir el tipo de intervención adecuado.

Consecuencias de la aplicación de tests

Braden y Niebling (2012) señalan que en la información sobre las consecuencias de la aplicación de un test debe incluirse la valoración individualizada de la utilidad diagnóstica. El diagnóstico o clasificación de un niño pequeño no debe basarse solo en los resultados de un único test, sino que también debe fundamentarse en la información recopilada mediante observación y otras modalidades de evaluación (p. ej., a través del juego) y la aplicación de distintos tests relevantes, así como en los datos proporcionados por otras fuentes (padres, maestros, etc.). Los resultados procedentes de estudios con grupos especiales y que tradicionalmente utilizan puntuaciones de corte indican el porcentaje de niños del grupo especial (p. ej., con alta capacidad intelectual, con discapacidad intelectual) que se sitúan en un rango de puntuación específico de la WPPSI-IV (p. ej., FSIQ ≥ 120 o FSIQ ≤ 75).

Braden and Niebling (2012) también afirman que, al analizar las consecuencias de la aplicación de tests, deben considerarse también las consecuencias inesperadas, como el sesgo de los ítems y las diferencias entre las puntuaciones obtenidas por los distintos grupos. El procedimiento seguido para analizar el sesgo de los ítems (funcionamiento diferencial de los ítems) y la adecuación del contenido se describe en el capítulo 3 y 4 de este manual. Las escalas de inteligencia de Wechsler han sido objeto de diversos estudios que han analizado las diferencias entre las puntuaciones obtenidas por los sujetos de distintas muestras, prestando especial atención a las agrupaciones por sexo y nivel socioeconómico (p. ej., Harris y Llorente, 2005; Weiss, Chen, Harris, Holdnack y Saklofske, 2010).

Los tests psicológicos y educativos se aplican, por lo general, con el convencimiento de que, a partir de los resultados obtenidos, los sujetos evaluados obtendrán algún beneficio. Además, los tests ayudan a los profesionales y a las instituciones a cumplir con los criterios legales y profesionales que conciernen al diagnóstico y clasificación, al plan de tratamiento o programa de rehabilitación, y a actividades de investigación.

Los profesionales deben tener siempre en cuenta las consecuencias esperadas e inesperadas del hecho de usar tests, así como las consecuencias esperadas e inesperadas de confiar en valoraciones informales en vez de en valoraciones basadas en los resultados de los tests. Aunque las consecuencias de la aplicación de tests pueden influir en la decisión sobre el uso de un test (si utilizarlo o no), las consecuencias adversas en sí mismas no disminuyen la validez de las interpretaciones del test previstas (AERA et al., 1999).

Interpretación de los resultados

Los resultados de la WPPSI-IV proporcionan información importante sobre el funcionamiento cognitivo de un niño, pero nunca deberían interpretarse aisladamente. Las respuestas a los ítems y las puntuaciones obtenidas en las pruebas ofrecen información cualitativa y cuantitativa que debe interpretarse junto con el historial detallado, acompañado de la observación clínica y minuciosa del niño. El profesional, además, suele disponer de otras fuentes de información: historial médico, educativo y psicosocial; observaciones conductuales directas; técnicas proyectivas; puntuaciones obtenidas en tests previos; aspectos cualitativos del rendimiento del niño en un test; y resultados obtenidos en otros instrumentos relevantes. Asimismo, el profesional debería evaluar los resultados en el contexto del motivo o del propósito de la evaluación.

Muchos de los métodos, estrategias y procedimientos de investigación desarrollados y perfeccionados por profesionales e investigadores para las anteriores escalas de inteligencia de Wechsler continúan siendo válidos y útiles (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008a, 2008b). No obstante, el profesional ha de tener en cuenta los cambios en el contenido de las distintas versiones cuando utilice estas directrices de interpretación.

Este capítulo se centra en los aspectos fundamentales de la interpretación de los resultados de la WPPSI-IV, especialmente en la presentación de las puntuaciones y el uso de las tablas a efectos interpretativos. Queda fuera del alcance de este capítulo proponer un método detallado, sistemático y completo de análisis e interpretación de los resultados; no obstante, en las siguientes secciones se exponen los principios básicos de interpretación, así como información sobre la presentación de los resultados y el análisis del perfil.

Descripción de los resultados

La conversión de las puntuaciones directas en puntuaciones escalares o compuestas (puntuaciones típicas) permite al profesional comparar las puntuaciones de las pruebas entre sí y con otras medidas relacionadas. El uso de puntuaciones típicas ajustadas a la edad permite comparar las aptitudes cognitivas de cada niño con las de otros niños de edad similar. Las puntuaciones típicas proporcionan la descripción más precisa sobre los resultados de un test; sin embargo, su interpretación puede resultar compleja para las personas que no están familiarizadas con los procedimientos de interpretación de tests. Para describir el rendimiento de un niño, a menudo se utilizan, junto con las puntuaciones típicas, otros datos, como los percentiles, los intervalos de confianza, la clasificación cualitativa y las edades equivalentes.

Puntuaciones típicas

La WPPSI-IV proporciona dos tipos de puntuaciones típicas ajustadas a la edad: puntuaciones escalares y puntuaciones compuestas. Las puntuaciones escalares representan el rendimiento de un niño en relación con los niños de su misma edad. Se calculan a partir de las puntuaciones directas de cada prueba y tienen una escala métrica con media 10 y desviación típica 3. Una puntuación escalar 10 refleja el rendimiento promedio de un determinado grupo de edad. Las puntuaciones escalares 7 y 13 se alejan una desviación típica por debajo y por encima de la media, respectivamente. Las puntuaciones escalares 4 y 16 se alejan dos desviaciones típicas por debajo y por encima de la media, respectivamente. Las puntuaciones escalares correspondientes a las

puntuaciones de procesamiento (CAA y CAE) se obtienen de la misma manera que las escalares de las pruebas. La tabla 7.1 presenta la relación entre las puntuaciones escalares, las desviaciones típicas respecto a la media y los percentiles equivalentes.

Tabla 7.1. Relación entre las puntuaciones escalares, las desviaciones típicas respecto a la media y los percentiles equivalentes

Puntuación escalar	*Dt* a la media	Percentil[a]
19	+3	99.9
18	$+2\frac{2}{3}$	99.6
17	$+2\frac{1}{3}$	99
16	+2	98
15	$+1\frac{2}{3}$	95
14	$+1\frac{1}{3}$	91
13	+1	84
12	$+\frac{2}{3}$	75
11	$+\frac{1}{3}$	63
10	0 (Media)	50
9	$-\frac{1}{3}$	37
8	$-\frac{2}{3}$	25
7	-1	16
6	$-1\frac{1}{3}$	9
5	$-1\frac{2}{3}$	5
4	-2	2
3	$-2\frac{1}{3}$	1
2	$-2\frac{2}{3}$	0.4
1	-3	0.1

[a] Los percentiles son valores teóricos para una distribución normal.

Las puntuaciones compuestas (ICV, IVE, IRF, IMT, IVP, CI total, IAV, INV, ICG e ICC) son puntuaciones típicas basadas en la suma de las puntuaciones escalares de varias pruebas. Las puntuaciones compuestas tienen una media 100 y una desviación típica 15. Una puntuación 100 en cualquiera de las puntuaciones compuestas define el rendimiento medio de los niños de la misma edad, y las puntuaciones 85 y 115 se corresponden con un alejamiento de una desviación típica por debajo y por encima de la media, respectivamente. Las puntuaciones 70 y 130 se alejan dos desviaciones típicas por debajo y por encima de la media, respectivamente. Aproximadamente el 68 % de todos los niños obtienen puntuaciones compuestas entre 85 y 115, alrededor del 96 % de los niños logran puntuaciones compuestas entre 70 y 130, y prácticamente todos los niños (alrededor del 99.8 %) obtienen puntuaciones entre 55 y 145 (que se alejan tres desviaciones típicas por debajo y por encima de la media, respectivamente). La tabla 7.2 presenta la relación entre las puntuaciones compuestas, las desviaciones típicas respecto a la media y los percentiles equivalentes.

Tabla 7.2. Relación entre las puntuaciones compuestas, las desviaciones típicas respecto a la media y los percentiles equivalentes

Puntuación compuesta	*Dt* a la media	Percentil[a]
160	+4	>99.9
155	$+3\frac{2}{3}$	>99.9
150	$+3\frac{1}{3}$	>99.9
145	+3	99.9
140	$+2\frac{2}{3}$	99.6
135	$+2\frac{1}{3}$	99
130	+2	98
125	$+1\frac{2}{3}$	95
120	$+1\frac{1}{3}$	91
115	+1	84
110	$+\frac{2}{3}$	75
105	$+\frac{1}{3}$	63
100	0 (Media)	50
95	$-\frac{1}{3}$	37
90	$-\frac{2}{3}$	25
85	-1	16
80	$-1\frac{1}{3}$	9
75	$-1\frac{2}{3}$	5
70	-2	2
65	$-2\frac{1}{3}$	1
60	$-2\frac{2}{3}$	0.4
55	-3	0.1
50	$-3\frac{1}{3}$	<0,1
45	$-3\frac{2}{3}$	<0,1
40	-4	<0,1

[a] Los percentiles son valores teóricos para una distribución normal.

Percentiles

Los percentiles indican la posición del niño en relación con otros niños de la misma edad. Los percentiles representan puntos de una escala bajo los cuales se sitúa un porcentaje concreto de puntuaciones, calculados a partir de la muestra de tipificación. Los percentiles oscilan normalmente entre 1 y 99, con una media y mediana de 50. Por ejemplo, un niño con un percentil 15 ha obtenido una puntuación igual o superior al 15 % de los niños de la misma edad (o dicho de otro modo, igual o inferior al 85 % del resto de los niños de la misma edad).

Aunque los percentiles son fáciles de entender y útiles para explicar el rendimiento de un niño en relación con el rendimiento de otros, tienen algunas limitaciones. La escala de percentiles no está

compuesta por intervalos iguales. En una distribución normal, como la del CI total, los percentiles tienden a agruparse en torno a la mediana (percentil 50). Por consiguiente, para los niños que obtienen puntuaciones alrededor de la media, una pequeña diferencia en su puntuación directa, apenas 1 o 2 puntos, puede dar a lugar a una gran diferencia en los percentiles; mientras que, para los niños con puntuaciones más extremas, el cambio de 1 o 2 puntos en las puntuaciones directas no dará lugar a una diferencia considerable en los percentiles.

Errores típicos de medida e intervalos de confianza

Las puntuaciones obtenidas en un test de aptitud cognitiva como la WPPSI-IV proporcionan una estimación de la puntuación verdadera del niño. Las puntuaciones verdaderas reflejan la aptitud real de un niño junto con cierto grado de error de medida. Para precisar al máximo la puntuación verdadera de un niño, es más adecuado establecer un intervalo de confianza (un rango de puntuaciones) en el que es probable que se sitúe su puntuación verdadera. Los intervalos de confianza proporcionan otras maneras de expresar la precisión de la puntuación y sirven como recordatorio de que el error de medida es inherente a todas las puntuaciones de los tests. Las puntuaciones menos fiables tienen un intervalo más amplio, mientras que las puntuaciones que son más fiables y precisas tienen un intervalo más pequeño. Se anima a los profesionales a registrar los intervalos de confianza de las puntuaciones compuestas y a usarlos para garantizar una mayor precisión en la interpretación de las puntuaciones. En el capítulo 5 de este manual pueden consultarse las fórmulas que permiten calcular los distintos intervalos de confianza de las puntuaciones de las pruebas y de las puntuaciones compuestas.

Las tablas A.2-A.11 del *Manual de aplicación y corrección* y las tablas B.1-B.3 y C.1-C.3 de este manual presentan los intervalos de confianza de las puntuaciones compuestas utilizando el error típico de estimación (*ETE*) y la puntuación verdadera estimada (*PVE*). Los intervalos de confianza calculados a partir del error típico de medida (*ETM*) se agrupan alrededor de las puntuaciones observadas, mientras que los intervalos de confianza calculados con el error típico de estimación (*ETE*) se agrupan alrededor de las puntuaciones verdaderas estimadas corregidas mediante regresión a la media.

Clasificación cualitativa

A menudo las puntuaciones compuestas se describen en términos cualitativos según el nivel de rendimiento de un niño. Las clasificaciones cualitativas proporcionan términos cualitativos que caracterizan el nivel de rendimiento de un niño respecto al de los niños de la misma edad. En la tabla 7.3. se presentan la clasificación cualitativa de las puntuaciones compuestas. Los resultados de los índices del test pueden describirse de una forma similar a como se describe el CI total en este ejemplo:

> En comparación con los niños de su misma edad, la puntuación obtenida por este niño en una prueba estandarizada de la aptitud intelectual caracteriza su funcionamiento intelectual en el rango [*insertar la clasificación cualitativa correspondiente*].

Tabla 7.3. Clasificación cualitativa de las puntuaciones compuestas

		Porcentajes	
Puntuación compuesta	**Clasificación cualitativa**	**Curva normal teórica**	**Muestra[a]**
130 y superior	Muy superior	2.5	2.24
120-129	Superior	7.2	8.14
110-119	Medio-alto	16.6	16.72
90-109	Medio	49.5	50.72
80-89	Medio- bajo	15.6	14.94
70-79	Límite	6.5	4.79
69 e inferior	Muy bajo	2.1	2.45

[a] Los porcentajes mostrados son para el CI total y se basan en la muestra de tipificación española (N = 885). Los porcentajes obtenidos para los índices son muy similares.

Puntuaciones de edad equivalente

La puntuación de edad equivalente representa la edad promedio, en años y meses, en la que es más frecuente obtener una determinada puntuación directa. Por ejemplo, una puntuación directa de 22 en la prueba *Cubos* se corresponde con una edad equivalente de 5:4 años, puesto que es la puntuación media obtenida por los niños de 5 años 4 meses. Las puntuaciones de edad equivalente que figuran en la tabla A.13 del *Manual de aplicación y corrección* son útiles para describir la aptitud cognitiva de un niño en comparación con el rendimiento promedio o típico de niños de distintas edades, pero debe emplearse con cierta cautela. En primer lugar, las puntuaciones de edad equivalente ofrecen poca información sobre la posición de un niño respecto a los niños de la misma edad. Un niño cuya edad equivalente es aparentemente significativa puede, o no, situarse en el rango promedio respecto a los niños de la misma edad. Por ejemplo, el niño A, de 6:0 años, ha obtenido una puntuación de edad equivalente de 4:10 años en la prueba *Reconocimiento*. Aunque puede parecer que el rendimiento del niño A en esta prueba es bajo, su puntuación directa equivale a una puntuación escalar de 8, que se encuentra en el rango promedio del rendimiento en comparación con los niños de la misma edad.

En segundo lugar, dado que las puntuaciones de edad equivalente se calculan a partir de las puntuaciones directas, no producen unidades o intervalos equivalentes en toda la escala. Por ello, los pequeños cambios de la puntuación directa pueden convertirse en grandes cambios en las puntuaciones de edad equivalente. Se pueden presentar importantes diferencias entre la edad equivalente y la edad cronológica de un niño, pero es inapropiado interpretar esta diferencia como reflejo de unas aptitudes cognitivas inferiores o superiores a la media. Esto sucede porque dentro del rango de puntuaciones medias se solapan diversos grupos de edad adyacentes. Por ejemplo, el niño B y el niño C tienen 5 años 7 meses, y a ambos se les aplicó la prueba *Nombres*. El niño B consiguió una puntuación directa de 18 y una edad equivalente de 5 años 4 meses (5:4). El niño C consiguió una puntuación directa de 17 y una edad equivalente de 4 años 10 meses (4:10). Esto no significa que la aptitud del niño B en *Nombres* sea 6 meses más avanzada que la del niño C. De hecho, la puntuación escalar del niño B, que es de 10, y la puntuación escalar del niño C, que es de 9, se encuentran en el mismo rango promedio respecto a los niños de la misma edad.

En tercer lugar, las puntuaciones de edad equivalente no son comparables entre las pruebas. Los percentiles correspondientes a dos pruebas con las mismas puntuaciones de edad equivalente pueden ser considerablemente distintos. Por ejemplo, si el niño D, de 4 años 10 meses, obtiene una puntuación de edad equivalente de 7:6 tanto en *Matrices* (puntuación directa de 19 y puntuación escalar de 15) como en *Búsqueda de animales* (puntuación directa de 52 y puntuación escalar de 18), los percentiles de estas pruebas son 95 y 99.6, respectivamente.

Finalmente, una puntuación de edad equivalente extrema no significa que las aptitudes cognitivas de un niño se asemejen a las del grupo de menor o mayor edad. Las puntuaciones equivalentes en los extremos de edad son más difíciles de interpretar porque de las puntuaciones extremas solo puede decirse que están por debajo de 2 años 6 meses o por debajo de 4 años 0 meses (en función de la franja de edad), o por encima de 7 años 6 meses.

Dadas estas limitaciones, se recomienda no utilizar las puntuaciones de edad equivalente como datos principales e interpretarlas con cautela. Para comparar el rendimiento de un niño respecto a los niños de la misma edad, deben usarse las puntuaciones típicas y los percentiles. Las decisiones clínicas han de basarse en el análisis de las puntuaciones típicas del niño, así como en informaciones cualitativas y en su historial. El diagnóstico y la toma de decisiones nunca deben basarse únicamente en las puntuaciones de edad equivalente, ni en ningún tipo de puntuación que se presente aislado.

Etapas para el análisis de un perfil

El rendimiento de un niño puede interpretarse en función de su perfil de puntuaciones escalares y puntuaciones compuestas. El análisis de un perfil puede realizarse desde una perspectiva intraindividual (comparando las puntuaciones compuestas y las puntuaciones de las pruebas entre sí) e interindividual (comparando el patrón de puntuaciones obtenidas con las del grupo normativo

correspondiente). Estas comparaciones aptitudinales pueden ayudar al profesional a identificar patrones potencialmente significativos de puntos fuertes y débiles, lo que es importante para describir trastornos funcionales y para diseñar y preparar planes de tratamiento (Lichtenberger y Kaufman, 2003; Sattler, 2008a, 2008b).

Un buen análisis del perfil del niño debe empezar con la generación de hipótesis basadas en el patrón de las puntuaciones obtenidas. Posteriormente, estas hipótesis se corroborarán o rechazarán en función de los resultados de otras evaluaciones, de su historial o de observaciones directas del comportamiento (Lichtenberger y Kaufman, 2003; Prifitera, Saklofske y Weiss, 2008). Aunque el cuadernillo de anotación se ha diseñado para ayudar al profesional en el análisis del perfil, los profesionales siempre deberían obtener cuantos más datos confirmatorios sobre el niño sea posible, incluidos los relativos a su nacimiento y desarrollo; antecedentes familiares y culturales; historial médico, educativo y social; y resultados obtenidos en evaluaciones anteriores. Toda esta información debe tenerse en cuenta al interpretar los resultados de un niño, y debe constar en los informes sobre el rendimiento del niño en la escala.

En la siguiente sección se describen los pasos básicos para realizar el análisis primario y secundario del perfil; no obstante, para llevar a cabo una interpretación más racional, es necesario que un profesional experto integre los datos obtenidos en el cuadro clínico del niño. Si se desea información más detallada sobre la interpretación clínica de las escalas de inteligencia de Wechsler, véanse Flanagan y Kaufman (2009), Lichtenberger y Kaufman (2003), Prifitera et al. (2008), y Sattler (2008a, 2008b).

Realización del análisis primario de un perfil

Paso 1. Descripción del CI total

El CI total es la puntuación más fiable y, tradicionalmente, la primera que se considera al interpretar un perfil. Es el resultado de la suma de cinco puntuaciones escalares de las pruebas para las edades de 2:6 a 3:11 años, y de seis puntuaciones escalares de las pruebas para las edades de 4:0 a 7:7 años. Por lo general, se considera que el CI total es la puntuación más representativa del funcionamiento de la aptitud intelectual general (*g*). Debe registrarse el CI total en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Para interpretar mejor el CI total, deben tenerse en cuenta los índices primarios y comparar las diferencias entre los índices primarios y otras puntuaciones. Los constructos medidos por los índices primarios suponen una parte importante de la inteligencia general. Se recomienda analizar los índices primarios en la primera fase de la interpretación clínica, en especial en los casos en que se da una variabilidad considerable entre los índices primarios y/o las puntuaciones de las pruebas.

Cada índice primario puede compararse con una estimación del rendimiento general (la MIP o el CI total); los índices primarios también pueden compararse entre ellos (p. ej., ICV–IRF). La variabilidad entre las puntuaciones escalares de las pruebas puede aportar más información relevante para interpretar el CI total. Así pues, para disponer de más datos relevantes para llevar a cabo la interpretación de los resultados, pueden compararse las puntuaciones escalares de las pruebas con una estimación del rendimiento general (la MPE-IP o la MPE-CI), pueden compararse las puntuaciones escalares de las pruebas entre ellas (p. ej., *Información–Semejanzas*), y puede realizarse el análisis de dispersión de las pruebas . Véanse Lichtenberger y Kaufman (2003) and Sattler (2008a) para información más detallada sobre la dispersión de las pruebas en la interpretación de las puntuaciones compuestas.

Paso 2. Descripción del ICV

El ICV es una medida del conocimiento adquirido a partir del entorno, la formación de conceptos verbales y el razonamiento verbal del niño. Todos los ítems de las pruebas de esta escala, incluso los ítems gráficos, se presentan verbalmente; y el niño debe dar respuestas verbales a todos los ítems verbales. Las pruebas que conforman el ICV de la WPPSI-IV para las edades de 2:6 a 3:11 años son las mismas que las que conforman el CI verbal de la WPPSI-III para la franja de menor edad. Sin embargo, las pruebas que conforman el ICV de la WPPSI-IV para las edades de 4:0 a

7:7 años son distintas de las que conforman el CI verbal de la WPPSI-III para la franja de mayor edad. *Información* es una prueba común del ICV y del CI verbal de la WPPSI-III. El ICV de la WPPSI-IV no incluye ni *Vocabulario* ni *Adivinanzas*, que forman parte del CI verbal de la WPPSI-III para la franja de mayor edad, y sí incluye *Semejanzas*, que no forma parte del CI verbal de la WPPSI-III. Por ello, el ICV de la WPPSI-IV puede considerarse una medida más precisa y pura del conocimiento adquirido, la formación de conceptos verbales y el razonamiento verbal, menos enfocada al conocimiento léxico y la expresión verbal que el CI verbal de la WPPSI-III. Debe registrarse la puntuación del ICV en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Paso 3. Descripción del IVE

El IVE es una medida del procesamiento visoespacial, de la integración y síntesis de las relaciones parte-todo, de la atención a los detalles visuales, de la formación de conceptos no verbales y de la integración visomotora. Las pruebas que conforman el IVE para las edades de 2:6 a 3:11 años son las mismas que las que conforman el CI manipulativo de la WPPSI-III para la franja de menor edad. Sin embargo, las pruebas que conforman el IVE de la WPPSI-IV para las edades de 4:0 a 7:7 años son distintas de las que conforman el CI manipulativo de la WPPSI-III. *Cubos* es una prueba común del IVE y del CI manipulativo de la WPPSI-III para la franja de mayor edad. El IVE de la WPPSI-IV no incluye ni *Matrices* ni *Conceptos*, que forman parte del CI manipulativo de la WPPSI-III, y sí incluye *Rompecabezas*, que no forma parte del CI manipulativo de la WPPSI-III. Por ello, respecto al CI manipulativo de la WPPSI-III para las edades de 4:0 a 7:7 años, el IVE incide más en la integración de las relaciones parte-todo, el procesamiento espacial, la integración visomotora y el rendimiento con limitación de tiempo. Debe registrarse el IVE en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Paso 4. Descripción del IRF (solo edades 4:0-7:7)

El IRF es una medida del razonamiento fluido e inductivo, la inteligencia visual general, el procesamiento simultáneo, el pensamiento conceptual y la clasificación. Las pruebas que conforman el IRF de la WPPSI-IV son distintas de las que conforman el CI manipulativo de la WPPSI-III. Tienen en común *Matrices* y *Conceptos*, pero no *Cubos*. Respecto al CI manipulativo de la WPPSI-III para las edades de 4:0 a 7:7 años, el IRF incide más en el razonamiento fluido e inductivo y en la clasificación, y menos en la integración de las relaciones parte-todo, el procesamiento espacial, la integración visomotora y el rendimiento con limitación de tiempo. Debe registrarse el IRF en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Paso 5. Descripción del IMT

El IMT es una medida de la memoria de trabajo visual, la memoria de trabajo visoespacial y la aptitud para mostrar resistencia a la interferencia proactiva. La memoria de trabajo se relaciona con la atención, la concentración, el control mental y el razonamiento. La investigación actual indica que la memoria de trabajo es un componente esencial de otros procesos cognitivos de nivel superior (Burgess y Braver, 2010; Hornung et al., 2011; Martínez et al., 2011). El IMT es un índice nuevo en la WPPSI-IV, al igual que las dos pruebas que lo conforman (*Reconocimiento* y *Localización*). Debe registrarse el IMT en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Paso 6. Descripción del IVP (solo edades 4:0-7:7)

El IVP es una medida de la aptitud del niño para rápidamente explorar, ordenar o discriminar información visual simple. Este índice también mide la memoria visual a corto plazo, la coordinación visomotora, la flexibilidad cognitiva, la discriminación visual, la concentración y la velocidad al realizar las pruebas (Groth-Marnat, 2003; Lichtenberger y Kaufman, 2003; Sattler, 2008a, 2008b). Diversos estudios señalan una correlación significativa entre la velocidad de procesamiento y la aptitud cognitiva (Dodonova y Dodonov, 2012), y subrayan la sensibilidad de las pruebas de velocidad de procesamiento a ciertos estados clínicos, como el TDAH, los trastornos del aprendizaje y los trastornos del espectro autista (Calhoun y Mayes, 2005; Hale

et al., 2012; Mayes et al., 2012). Es importante señalar que las pruebas que conforman el IVP no son solo medidas del tiempo de reacción o de la discriminación visual, puesto que en ellas también intervienen los procesos cognitivos de toma de decisiones o los procesos de aprendizaje.

Las pruebas de Velocidad de procesamiento (VP) de la WPPSI-III son *Claves* y *Búsqueda de símbolos;* mientras que las pruebas que conforman el IVP de la WPPSI-IV son *Búsqueda de animales* y *Cancelación*. Respecto al índice VP de la WPPSI-III, los estímulos de las pruebas del IVP son más conocidos y sugerentes para los niños pequeños, los requisitos cognitivos son más adecuados al desarrollo, y en las respuestas interviene menos la coordinación motriz fina. Por ello, el IVP de la WPPSI-IV puede considerarse una medida más precisa y pura de la velocidad de procesamiento, más adecuada al desarrollo y menos dependiente de las destrezas de la motricidad fina que el índice VP de la WPPSI-III.

Al igual que con todas las puntuaciones compuestas, la interpretación del IVP será más acertada cuanto más similares sean las puntuaciones obtenidas en las pruebas que contribuyen al índice. El profesional debe determinar si el funcionamiento motor del sujeto puede afectar de algún modo a la puntuación del IVP. Cuando el IVP es bajo, en relación con el MIP o el CI total, el profesional debe considerar si la velocidad del niño al responder puede afectar a su rendimiento en los tests. Debe registrarse el IVP en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Paso 7. Evaluación de los puntos fuertes y débiles en los índices

Es inusual que un niño normal presente el mismo rendimiento en todas las aptitudes. Para obtener información sobre los puntos fuertes y débiles cognitivos de un niño, respecto a su nivel de aptitud cognitiva general, pueden compararse los índices primarios con un indicador del rendimiento general (la MIP o el CI total).

La diferencia entre un índice primario y un indicador del rendimiento general puede ser clínicamente significativo para un niño, pero no para otro. La interpretación de las diferencias debe corroborarse mediante otras evidencias, y el uso de otros instrumentos de evaluación puede servir para refutar o confirmar una hipótesis. Si los datos obtenidos no corroboran la hipótesis original y apuntan a otras hipótesis, estas pueden evaluarse mediante otros instrumentos u otras técnicas que permiten obtener información. Por ejemplo, si un niño muestra un rendimiento significativamente más bajo en el IVE respecto a la MIP, la baja puntuación puede atribuirse a que su aptitud visoespacial es baja o sus destrezas motoras son insuficientes. En este caso, las evaluaciones adicionales podrían centrarse en tareas de aptitud visoespacial sin requisitos motores, o bien en tareas motoras en las que no interviniera la aptitud visoespacial. Para confirmar o rechazar las hipótesis, deben tenerse en cuenta la información recopilada mediante entrevistas, las observaciones directas, así como los resultados de otras evaluaciones.

La interpretación de los puntos fuertes o débiles reflejados en los índices dependen, en parte, del indicador usado en la comparación (la MIP o el CI total). La proporción de pruebas de cada escala primaria es distinta en la MIP y el CI total: algunas pruebas que contribuyen a la MIP no contribuyen al CI total. Por lo tanto, el contenido de estos indicadores es ligeramente distinto, y las pruebas que los componen deben tenerse en cuenta al realizar la interpretación.

En la página de análisis primario del cuadernillo de anotación se dispone de una sección para identificar los puntos fuertes y débiles del niño en los índices. El profesional ha de elegir la MIP o el CI total como puntuación de comparación. Si es posible, es preferible utilizar la MIP porque se basa en más pruebas y es una muestra más amplia del rendimiento que el CI total. Asimismo, el profesional debe elegir un nivel de significación de los valores críticos. En la elección de este nivel de significación interviene el juicio clínico. La decisión tomada debe basarse en el tipo y número de comparaciones que vayan a realizarse, el objetivo de la evaluación y el motivo de consulta. En el apartado «Opciones de comparación» del capítulo 2 del *Manual de aplicación y corrección* y en el capítulo 5 de este manual se exponen más consideraciones psicométricas y clínicas sobre la elección del nivel de significación más adecuado.

La tabla B.1 del *Manual de aplicación y corrección* presenta los valores críticos corregidos para realizar comparaciones entre cada uno de los índices primarios y la MIP o el CI total con los niveles de significación .01, .05, .10 y .15. El profesional puede usar esta tabla para decidir si las diferencias entre puntuaciones son significativas estadísticamente. Cabe señalar que los valores se basan en la aplicación de las pruebas principales que conforman cada puntuación. Los valores de esta tabla han de usarse con cautela al valorar los puntos fuertes y débiles de un niño, si se ha realizado alguna sustitución de las permitidas; dado que es posible que se haya introducido un error de medida adicional, debe registrarse en el informe.

En la tabla B.2 del *Manual de aplicación y corrección* figuran las tasas base de las diferencias entre cada uno de los índices primarios y la MIP o el CI total. Estas tasas base se presentan en las dos direcciones de la diferencia, para la muestra total y según el nivel de aptitud reflejado en el CI total. Con esta tabla, el profesional puede determinar la frecuencia de las diferencias respecto a la muestra de tipificación. Los profesionales a menudo se preguntan qué nivel de ocurrencia (tasa base) debe considerarse poco frecuente: 15 %, 10 %, 5 % o 1 %. Para determinar si la diferencia es poco frecuente, deben tenerse en cuenta, además del juicio clínico, otros aspectos, como el desarrollo lingüístico, motor y cognitivo del niño, y su entorno cultural. Sattler (2008a) sugiere que si las diferencias entre puntuaciones se dan en menos del 15 % de la muestra de tipificación, deben considerarse inusuales.

Paso 8. Evaluación de las comparaciones entre los índices

Wechsler (1944) destacó el CI total y lo describió como el promedio del rendimiento mostrado por un sujeto en todas las pruebas. Sin embargo, Wechsler observó que a veces era conveniente analizar por separado las puntuaciones de pruebas verbales y manipulativas, en especial en personas «con trastornos especiales que requieren una atención especial» (Wechsler, 1994, p. 138). Aunque es más probable que se hallen diferencias estadísticamente significativas entre las puntuaciones de los índices primarios en poblaciones clínicas, con frecuencia también se encuentran diferencias similares en la muestra de tipificación (Lichtenberger y Kaufman, 2003; Sattler, 2008a).

Los profesionales deben tener en cuenta que las puntuaciones de los índices primarios son estimaciones del funcionamiento general del niño en un dominio cognitivo o área de contenido particular. Por ello, estas puntuaciones siempre deben analizarse en el contexto de las dos pruebas que las conforman. Una discrepancia extrema entre las dos puntuaciones escalares de las pruebas que contribuyen a cada índice primario indica que la puntuación de dicho índice representa la síntesis de diversas aptitudes. Al interpretar las puntuaciones de los índices primarios, el profesional debe examinar minuciosamente los resultados obtenidos por el niño en las dos pruebas que conforman cada índice. Cuando se observa una diferencia significativa y poco frecuente, debe tenerse en cuenta esta variabilidad al interpretar y comparar las puntuaciones de los índices primarios (Sattler, 2008a).

En la página de análisis primario del cuadernillo de anotación se incluye una tabla que facilita el análisis de las diferencias entre índices. La tabla B.5 del *Manual de aplicación y corrección* ofrece los valores críticos de las diferencias estadísticamente significativas entre índices primarios con los niveles de significación .01, .05, .10 y .15, por grupo y franja de edad. Si las puntuaciones de los índices primarios no son significativamente diferentes, los niveles de aptitud del niño en los dos dominios cognitivos que se comparan son bastante similares. Si las puntuaciones de los índices primarios son significativamente diferentes, es probable que los niveles de aptitud del niño en los dos dominios cognitivos que se comparan sean distintos. En la tabla B.6 del *Manual de aplicación y corrección* figuran las tasas base de las diferencias entre las puntuaciones de los índices primarios obtenidas por la muestra de tipificación. Estas tasas base se presentan en las dos direcciones de la diferencia, para la muestra total y según el nivel de aptitud reflejado en el CI total.

Paso 9. Evaluación de los puntos fuertes y débiles en las pruebas

Al igual que se ha visto respecto a los índices primarios, la mayoría de los niños presentan puntos débiles y fuertes en las pruebas relacionados con su rendimiento en las distintas aptitudes cognitivas. Al analizar estos puntos fuertes y débiles, debe seguirse un criterio claro. El significado clínico de estos resultados debe contextualizarse con la información recopilada y con las puntuaciones de las pruebas, y nunca debe confirmarse o rechazarse una hipótesis en función de una única observación.

Cuando se compara la puntuación de una prueba con un indicador general del rendimiento, la interpretación depende, en parte, de la contribución de la prueba en el indicador. Las puntuaciones de las pruebas que contribuyen a los índices primarios pueden compararse con la MPE-IP o la MPE-CI. Los profesionales deben considerar qué pruebas contribuyen al indicador de comparación al interpretar los puntos fuertes y débiles de un niño en las pruebas.

En la página de análisis primario del cuadernillo de anotación se dispone de una sección para identificar los puntos fuertes y débiles del niño en las pruebas. El profesional ha de elegir la MPE-IP o la MPE-CI total como puntuación de comparación. Si es posible, es preferible utilizar la MPE-IP porque se basa en más pruebas y es una muestra más amplia del rendimiento que la MPE-CI. Asimismo, el profesional debe elegir un nivel de significación de los valores críticos.

La tabla B.3 del *Manual de aplicación y corrección* presenta los valores críticos para realizar comparaciones entre cada una de las puntuaciones escalares de las pruebas que conforman los índices primarios y la MPE-IP o la MPE-CI con los niveles de significación .01, .05, .10 y .15, por grupo y franja de edad. El profesional puede usar esta tabla para decidir si las diferencias entre puntuaciones son significativas estadísticamente. Dado que los valores para realizar comparaciones con la MPE-CI, que figuran en la tabla B.3, se basan en las pruebas principales para obtener el CI total, si se utiliza la MPE-CI como puntuación de comparación con las pruebas, debe comprobarse si se ha realizado alguna sustitución, de las permitidas, para obtener el CI total. En tal caso, debe considerarse el valor crítico y la tasa base de la prueba principal en vez del valor crítico y la tasa base de la prueba sustituida. Por ejemplo, si *Rompecabezas* ha sustituido a *Cubos* para obtener el CI total (y, por tanto, contribuyó a la MPE-CI), debe considerarse el valor crítico y la tasa base de *Cubos* (la prueba principal que fue sustituida) en vez del valor crítico y la tasa base de *Rompecabezas*. Se sigue este proceso aparentemente contradictorio porque los valores críticos y las tasas bases de la prueba principal (p. ej., *Cubos*) tienen en cuenta el contenido compartido entre las puntuaciones, mientras que los valores de la prueba complementaria (p. ej., *Rompecabezas*) no los tienen en cuenta. En la tabla B.4 del *Manual de aplicación y corrección* figuran las tasas base de las diferencias entre las pruebas que conforman los índices primarios y la MPE-IP o la MPE-CI. Con esta tabla, el profesional puede determinar la frecuencia de las diferencias respecto a la muestra de tipificación.

Paso 10. Evaluación de las comparaciones entre las pruebas

Dado que cada índice está formado por dos pruebas, la comparación de las puntuaciones escalares de estas dos pruebas es informativa. Por ejemplo, la comparación de las puntuaciones obtenidas en *Reconocimiento* y *Localización* puede servir para interpretar el IMT y aportar información relevante sobre el rendimiento en la memoria visual, la memoria visoespacial y la memoria de trabajo. En la página de análisis del cuadernillo de anotación se incluye una tabla en la que realizar las comparaciones entre cada par de pruebas que contribuyen a cada índice primario.

La tabla B.7 del *Manual de aplicación y corrección* presenta los valores críticos de las diferencias estadísticamente significativas entre cada par de pruebas con los niveles de significación .01, .05, .10 y .15. Por lo general, debe seleccionarse un nivel de significación más estricto si se van a realizar todas las comparaciones, puesto que la probabilidad de errores tipo I (falsos positivos) es mayor cuando se comparan muchas puntuaciones al mismo tiempo. En el apartado «Opciones de comparación» del capítulo 2 del *Manual de aplicación y corrección* y en el capítulo 5 de este manual se ofrece información detallada sobre la elección del nivel de significación más adecuado. En la tabla B.8 del *Manual de aplicación y corrección* figuran las tasas base de las diferencias entre las puntuaciones escalares de las pruebas obtenidas por la muestra de tipificación. Estas tasas base se presentan en las dos direcciones de la diferencia.

Realización del análisis secundario de un perfil

En muchos casos no es necesario realizar el análisis secundario. Los pasos de interpretación que componen el análisis secundario son opcionales y, en función del motivo de evaluación, pueden ser necesarios únicamente algunos pasos. Generalmente, se realiza el análisis secundario debido a un objetivo clínico específico.

Paso 1. Descripción del IAV

El IAV ofrece información sobre el rendimiento de los niños con problemas del lenguaje expresivo. Es el resultado de la suma de las puntuaciones de *Dibujos* y *Nombres*, y es un indicador de la adquisición de vocabulario receptivo y expresivo del niño. En la prueba *Dibujos*, el niño debe señalar la respuesta correcta, mientras que en *Nombres* debe indicar oralmente su respuesta.

En el anexo B de este manual se ofrece más información sobre posibles usos del IAV, así como la tabla que permite obtener la puntuación de este índice secundario para ambas franjas de edad (tabla B.1). Debe registrarse el IAV en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Paso 2. Descripción del INV

El INV está conformado por pruebas que no requieren respuestas verbales. Es el resultado de la suma de cuatro puntuaciones escalares de las pruebas para las edades de 2:6 a 3:11 años, y de cinco puntuaciones escalares de las pruebas para las edades de 4:0 a 7:7 años. Este índice incluye pruebas de los dominios cognitivos primarios que contribuyen al CI total (p. ej., pruebas de Visoespacial y de Memoria de trabajo), excepto de Comprensión verbal. Dado que la intervención del lenguaje en las pruebas que conforman el INV es relativamente menor, este índice puede ofrecer una estimación más adecuada de la aptitud general que el CI total para los niños con retrasos del lenguaje expresivo o con trastornos clínicos asociados a problemas del lenguaje expresivo (p. ej., trastornos del espectro autista), así como niños que están aprendiendo español.

En el anexo B de este manual se ofrece más información sobre posibles usos del INV, así como las tablas que permiten obtener la puntuación de este índice secundario para cada franja de edad (tablas B.2 y B.3). Debe registrarse el INV en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Paso 3. Descripción del ICG

El ICG se basa en las pruebas de Comprensión verbal, Visoespacial y (solo para la franja de edad de 4:0 a 7:7 años) Razonamiento fluido que contribuyen al CI total. Conceptualmente, el ICG ofrece una estimación de la aptitud intelectual general menos dependiente de la memoria de trabajo y de la velocidad de procesamiento que el CI total.

En el anexo C de este manual se ofrece más información sobre posibles usos del ICG, así como las tablas que permiten obtener la puntuación de este índice secundario para cada franja de edad (tablas C.1 y C.2). Debe registrarse el ICG en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Paso 4. Descripción del ICC (solo edades 4:0-7:7)

El ICC es una puntuación compuesta que se basa en las pruebas que contribuyen al IMT y al IVP, y solo puede obtenerse para las edades de 4:0 a 7:7 años. Conceptualmente, el ICC ofrece una estimación de la eficiencia con la que se procesa la información cognitiva durante el aprendizaje, la resolución de problemas y el razonamiento de nivel superior.

En el anexo C de este manual se ofrece más información sobre posibles usos del ICC, así como la tabla que permite obtener la puntuación de este índice secundario para la franja de edad de 4:0 a 7:7 años (tabla C.3). Debe registrarse el ICC en el cuadernillo de anotación con el percentil y el intervalo de confianza correspondientes.

Paso 5. Evaluación de las comparaciones entre los índices

En la página de análisis secundario del cuadernillo de anotación se incluye una tabla que facilita el análisis de las diferencias entre los índices. El ICG puede compararse con el CI total para ambas franjas de edad, mientras que el ICC solo puede compararse para la franja de edad de 4:0 a 7:7 años.

La tabla C.4 de este manual ofrece los valores críticos de las diferencias estadísticamente significativas con los niveles de significación .01, .05, .10 y .15, por grupo y franja de edad. La tabla C.5 presenta las tasas base correspondientes a las diferencias entre el ICG y el ICC obtenidas por la muestra de tipificación. Estas tasas base se presentan en las dos direcciones de la diferencia, para la muestra total y según el nivel de aptitud reflejado en el ICG. En el anexo C de este manual se describen las posibles aplicaciones de las diferencias significativas y poco frecuentes entre el ICG y el CI total y entre el ICG y el ICC.

Paso 6. Evaluación de las comparaciones entre las pruebas de los índices secundarios

La comparación de las pruebas *Dibujos* y *Nombres* es relevante para interpretar el IAV porque permite al profesional determinar las diferencias entre la adquisición del vocabulario receptivo y expresivo, así como describir con más precisión la contribución de estas aptitudes al IAV.

La tabla B.7 del *Manual de aplicación y corrección* presenta los valores críticos de las diferencias estadísticamente significativas entre *Dibujos* y *Nombres* con los niveles de significación .01, .05, .10 y .15. En la tabla B.8 del *Manual de aplicación y corrección* figuran las tasas base de las diferencias entre las puntuaciones de *Dibujos* y *Nombres* obtenidas por la muestra de tipificación. Estas tasas base se presentan en las dos direcciones de la diferencia. En el anexo B de este manual se exponen las posibles aplicaciones de las diferencias significativas y poco frecuentes entre las puntuaciones de *Dibujos* y *Nombres*.

Paso 7. Evaluación de las comparaciones entre las puntuaciones de procesamiento (solo edades 4:0-7:7)

Las puntuaciones de procesamiento Cancelación aleatoria (CAA) y Cancelación estructurada (CAE) de la WPPSI-IV se crearon para ofrecer más información sobre las aptitudes cognitivas que influyen en el rendimiento del niño en la prueba *Cancelación*. Estas puntuaciones miden la atención selectiva y la velocidad de procesamiento del niño ante dos presentaciones distintas de los estímulos: aleatoria y estructurada. A diferencia de otras tareas de cancelación que requieren que el niño realice simplemente la correspondencia visoperceptiva entre los estímulos, la prueba *Cancelación* de la WPPSI-IV requiere que el niño determine si los estímulos pertenecen a una determinada categoría (prendas de vestir). Para discriminar los estímulos que ha de buscar y los distractores, el niño también puede realizar asociaciones visosemánticas.

La comparación de la diferencia entre las puntuaciones CAA y CAE proporciona información sobre la variación del rendimiento en *Cancelación* en función de la disposición de los estímulos visuales. Una puntuación superior en CAE respecto a CAA sugiere que al niño le favorece la presentación estructurada. Una puntuación superior en CAA respecto a CAE puede indicar que el niño aplica la estrategia de resolución de una forma más efectiva en la tarea CAA que en la CAE, o que al niño no le favorece la presentación estructurada (el niño puede haber utilizado una estrategia de búsqueda desorganizada a pesar de que la disposición de los estímulos visuales sea estructurada). Los datos obtenidos mediante la observación de las estrategias de búsqueda seguidas por el niño pueden aportar más información que permita ratificar o rechazar hipótesis relativas a la influencia que la presentación estructurada tiene sobre el rendimiento del niño en *Cancelación*.

En la página de análisis secundario del cuadernillo de anotación se incluye una tabla en la que realizar la comparación entre las puntuaciones CAA y CAE. La tabla B.7 del *Manual de aplicación y corrección* presenta los valores críticos de las diferencias estadísticamente significativas entre CAA y CAE con los niveles de significación .01, .05, .10 y .15. En la tabla B.8 del *Manual de aplicación y corrección* figuran las tasas base de las diferencias entre las puntuaciones de CAA y CAE obtenidas por la muestra de tipificación. Estas tasas base se presentan en las dos direcciones de la diferencia.

Realización del análisis cualitativo de las respuestas

El último paso del análisis del perfil es el análisis cualitativo de las respuestas del sujeto. Debe prestarse especial atención a las respuestas poco frecuentes y al patrón de respuestas general del niño. Las respuestas poco frecuentes a menudo ofrecen información muy reveladora. También es conveniente examinar el proceso mediante el cual el niño ha llegado a la respuesta correcta, o las razones por las que ha dado una respuesta incorrecta. El enfoque de procesamiento para la interpretación defendido por Kaplan y sus colegas (Kaplan, 1988; Kaplan, Fein, Morris y Delis, 1991) fue diseñado para ayudar al profesional a determinar la naturaleza de los errores que se cometían en los tests estandarizados. Aunque cada una de las pruebas de una escala de inteligencia se ha diseñado para medir un proceso cognitivo determinado, al realizar la prueba, suelen intervenir muchos procesos cognitivos. Son varias las razones por las que un niño puede obtener 0 puntos en un ítem. Por ejemplo, los niños pueden fallar un ítem de *Cubos* porque no perciben correctamente el diseño, no son capaces de analizar la configuración del diseño o agotan el tiempo concedido para su realización.

Resumen

La comparación y la interpretación de las puntuaciones que se registran en el cuadernillo de anotación de la WPPSI-IV deben llevarlas a cabo solo profesionales expertos. La evaluación psicológica es un acto clínico que utiliza las puntuaciones de los tests considerándolas únicamente como una fuente más de información que permite a un psicólogo con un buen juicio clínico hacer un retrato psicológico exhaustivo de un un niño. El profesional que utiliza los tests como parte de la evaluación es el responsable de la interpretación de los resultados y de la toma de decisiones relacionadas con el diagnóstico, el tratamiento o la intervención.

Anexo A

Tablas de intercorrelaciones

Tabla A.1. Intercorrelaciones de las puntuaciones de las pruebas y de las puntuaciones compuestas para las edades 2:6-2:11

Prueba/puntuación compuesta	I	D	N	C	RO	R	L	ICV	IVE	IMT	CIT	IAV	INV	ICG
Información								.44			.55			.52
Dibujos	.44							.44			.49	.35		.45
Nombres	.40	.35										.35		
Cubos	.38	.33	.30						.34		.48		.42	.47
Rompecabezas	.32	.24	.25	.34					.34		.44		.39	.40
Reconocimiento	.42	.42	.29	.35	.38					.57	.55		.56	
Localización	.36	.29	.13	.34	.24	.57				.57			.50	
ICV	.86	.83	.44	.42	.34	.49	.39							
IVE	.43	.36	.34	.85	.78	.44	.36	.47						
IMT	.42	.37	.23	.36	.33	.86	.91	.47	.43					
CIT	.74	.68	.46	.71	.65	.71	.52	.84	.83	.67				
IAV	.49	.81	.84	.36	.28	.41	.25	.76	.40	.36	.68			
INV	.49	.41	.34	.71	.65	.75	.75	.54	.84	.85	.88	.45		
ICG	.75	.69	.46	.75	.66	.55	.44	.85	.86	.53	.98	.68	.82	
Media	9.9	9.3	9.5	9.7	9.9	10.2	9.8	19.2	19.6	20.1	49.0	18.9	39.8	38.7
Dt	3.1	2.8	2.9	3.4	2.8	2.6	3.1	4.9	5.1	5.0	10.2	4.6	8.4	8.6

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Tabla A.2. Intercorrelaciones de las puntuaciones de las pruebas y de las puntuaciones compuestas para las edades 3:0-3:5

Prueba/puntuación compuesta	I	D	N	C	RO	R	L	ICV	IVE	IMT	CIT	IAV	INV	ICG
Información								.36			.52			.49
Dibujos	.36							.36			.42	.37		.36
Nombres	.51	.37										.37		
Cubos	.33	.26	.09						.31		.41		.34	.41
Rompecabezas	.35	.20	.21	.31					.31		.40		.43	.39
Reconocimiento	.38	.36	.36	.26	.27					.38	.46		.42	
Localización	.41	.35	.30	.17	.35	.38				.38			.42	
ICV	.80	.84	.53	.35	.33	.45	.46							
IVE	.42	.28	.19	.81	.81	.33	.32	.42						
IMT	.48	.43	.40	.26	.37	.84	.82	.55	.39					
CIT	.71	.66	.46	.65	.64	.68	.50	.83	.79	.71				
IAV	.52	.83	.82	.22	.25	.43	.39	.83	.28	.50	.68			
INV	.54	.42	.35	.65	.71	.70	.68	.58	.84	.83	.90	.47		
ICG	.73	.67	.43	.69	.67	.46	.46	.85	.84	.56	.96	.67	.84	
Media	10.4	10.7	10.1	9.8	10.0	10.4	9.9	21.2	19.8	20.3	51.3	20.8	40.1	40.9
Dt	2.9	3.2	3.1	3.0	3.0	3.0	2.8	5.0	4.9	4.8	10.1	5.2	8.1	8.4

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Tabla A.3. Intercorrelaciones de las puntuaciones de las pruebas y de las puntuaciones compuestas para las edades 3:6-3:11

Prueba/puntuación compuesta	I	D	N	C	RO	R	L	ICV	IVE	IMT	CIT	IAV	INV	ICG
Información								.53			.61			.59
Dibujos	.53							.53			.47	.52		.43
Nombres	.60	.52										.52		
Cubos	.42	.31	.38						.41		.52		.46	.50
Rompecabezas	.35	.18	.41	.41					.41		.47		.51	.40
Reconocimiento	.46	.40	.46	.42	.47					.51	.61		.63	
Localización	.35	.24	.41	.26	.30	.51				.51			.45	
ICV	.87	.88	.64	.41	.30	.50	.34							
IVE	.45	.28	.47	.83	.85	.53	.34	.42						
IMT	.46	.37	.50	.39	.44	.86	.88	.48	.50					
CIT	.77	.68	.66	.71	.68	.77	.46	.82	.82	.70				
IAV	.64	.87	.87	.39	.34	.49	.37	.87	.43	.50	.77			
INV	.53	.38	.56	.70	.74	.81	.71	.52	.86	.87	.88	.54		
ICG	.79	.70	.66	.73	.68	.61	.40	.85	.83	.58	.98	.78	.81	
Media	10.1	9.9	9.9	9.9	9.8	10.3	10.4	20.0	19.7	20.8	50.0	19.8	40.4	39.6
DI	3.0	3.1	3.1	2.9	3.1	2.9	3.2	5.3	5.0	5.3	10.7	5.3	8.9	8.7

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Tabla A.4. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 4:0-4:5

Prueba/puntuación de procesamiento/ puntuación compuesta	I	S	V	CO	D	N	C	RO	M	CON	R	L	BA	CA	CF	CAA	CAE	ICV	IVE	IRF	IMT	IVP	CIT	IAV	INV	ICG	ICC
Información																		.49					.60			.61	
Semejanzas	.49																	.49					.55			.50	
Vocabulario	.52	.62																									
Comprensión	.40	.58	.50																								
Dibujos	.33	.19	.36	.40																				.46			
Nombres	.58	.28	.50	.42	.46																			.46			
Cubos	.41	.27	.46	.24	.20	.34													.39				.43		.39	.43	
Rompecabezas	.28	.42	.44	.27	.25	.31	.39												.39								
Matrices	.48	.42	.49	.28	.24	.35	.39	.45												.25			.54		.49	.55	
Conceptos	.16	.39	.44	.34	.23	.18	.17	.51	.25											.25					.31		
Reconocimiento	.29	.25	.37	.25	.25	.27	.16	.29	.28	.15											.39		.35		.32		.41
Localización	.10	.36	.27	.20	.09	.10	.16	.27	.16	.28	.39										.39						.34
Búsqueda de animales	.36	.45	.48	.21	.32	.29	.35	.43	.36	.28	.31	.31										.59	.51		.49		.58
Cancelación	.40	.37	.45	.29	.31	.18	.26	.32	.29	.32	.24	.09	.59									.59					.42
Clave de figuras	.41	.43	.44	.23	.21	.27	.15	.23	.27	.20	.31	.20	.59	.49													
Cancelación aleatoria	.37	.41	.51	.35	.28	.16	.29	.34	.38	.32	.29	.14	.56	.91	.44												
Cancelación estructurada	.37	.26	.34	.21	.29	.14	.19	.23	.21	.29	.18	.06	.52	.93	.45	.72											
ICV	.84	.89	.66	.58	.29	.49	.38	.41	.52	.33	.31	.27	.47	.45	.49	.45	.35										
IVE	.41	.41	.54	.31	.27	.39	.82	.85	.50	.41	.27	.26	.47	.34	.23	.38	.25	.47									
IRF	.39	.51	.59	.40	.30	.33	.34	.61	.75	.83	.26	.29	.40	.39	.29	.44	.32	.53	.58								
IMT	.24	.36	.39	.27	.20	.23	.19	.34	.27	.26	.84	.83	.37	.20	.31	.26	.14	.35	.32	.33							
IVP	.43	.46	.52	.28	.36	.26	.34	.42	.36	.34	.31	.23	.89	.89	.61	.82	.81	.51	.46	.44	.32						
CIT	.74	.72	.72	.52	.37	.52	.62	.55	.71	.34	.56	.35	.68	.51	.51	.55	.42	.84	.70	.64	.55	.67					
IAV	.53	.27	.50	.48	.85	.85	.32	.33	.35	.24	.30	.11	.36	.28	.28	.26	.25	.45	.39	.37	.25	.36	.52				
INV	.52	.56	.70	.42	.39	.44	.62	.65	.70	.61	.57	.40	.71	.53	.47	.57	.44	.63	.76	.82	.59	.70	.91	.48			
ICG	.79	.74	.70	.53	.31	.52	.69	.51	.77	.32	.32	.25	.48	.42	.40	.46	.33	.88	.72	.66	.34	.51	.94	.49	.80		
ICC	.42	.52	.57	.34	.35	.30	.33	.47	.39	.37	.68	.62	.80	.70	.57	.69	.62	.54	.48	.48	.78	.84	.76	.38	.79	.54	
Media	9.8	10.0	9.7	10.0	10.1	10.4	10.3	10.3	10.0	9.6	10.0	10.0	10.0	10.0	9.8	10.1	10.0	19.8	20.6	19.5	20.0	20.0	60.3	20.4	49.8	40.2	40.0
Dt	2.6	3.0	3.0	3.2	2.5	2.5	2.8	3.0	3.0	3.5	2.8	2.7	3.0	2.9	2.8	3.1	3.0	4.8	4.9	5.1	4.6	5.3	11.4	4.3	9.7	8.4	8.1

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Tabla A.5. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 4:0-4:11

Prueba/puntuación de procesamiento/ puntuación compuesta	I	S	V	CO	D	N	C	RO	M	CON	R	L	BA	CA	CF	CAA	CAE	ICV	IVE	IRF	IMT	IVP	CIT	IAV	INV	ICG	ICC
Información																		.51					.55			.56	
Semejanzas	.51																	.51					.54			.56	
Vocabulario	.59	.62																									
Comprensión	.46	.36	.47																								
Dibujos	.65	.33	.43	.46																				.55			
Nombres	.67	.50	.49	.49	.55																			.55			
Cubos	.45	.36	.30	.29	.30	.33													.56				.56		.53	.50	
Rompecabezas	.40	.35	.31	.23	.45	.32	.56												.56								
Matrices	.34	.42	.23	.32	.34	.37	.37	.35												.37			.51		.47	.48	
Conceptos	.39	.40	.28	.24	.31	.40	.39	.31	.37											.37					.53		
Reconocimiento	.37	.41	.46	.28	.25	.29	.40	.35	.30	.35											.25		.52		.47		.38
Localización	.31	.24	.28	.34	.30	.30	.41	.19	.19	.32	.25										.25						.28
Búsqueda de animales	.25	.19	.19	.16	.25	.22	.35	.34	.34	.43	.35	.21										.45	.41		.51		.48
Cancelación	.13	.19	.18	.12	.18	.14	.25	.29	.33	.28	.22	.18	.45									.45					.39
Clave de figuras	.19	.08	.25	.27	.18	.28	.35	.20	.42	.38	.39	.42	.48	.49													
Cancelación aleatoria	.09	.21	.16	.07	.12	.12	.23	.28	.30	.26	.22	.16	.41	.94	.49												
Cancelación estructurada	.16	.16	.18	.17	.23	.18	.25	.27	.34	.30	.21	.19	.45	.95	.48	.81											
ICV	.86	.88	.70	.47	.56	.67	.46	.43	.44	.46	.45	.32	.25	.18	.15	.17	.18										
IVE	.48	.40	.35	.29	.43	.37	.87	.90	.41	.40	.42	.33	.39	.31	.31	.29	.29	.51									
IRF	.45	.49	.31	.34	.39	.46	.46	.40	.83	.82	.39	.31	.47	.37	.49	.34	.39	.54	.49								
IMT	.43	.41	.46	.40	.35	.38	.51	.34	.31	.42	.77	.80	.35	.25	.51	.24	.25	.48	.48	.44							
IVP	.22	.22	.21	.17	.25	.21	.35	.37	.39	.42	.33	.23	.85	.85	.57	.79	.83	.25	.41	.49	.35						
CIT	.71	.71	.58	.46	.52	.58	.70	.57	.69	.57	.69	.39	.60	.38	.47	.36	.39	.81	.72	.76	.67	.58					
IAV	.75	.47	.53	.54	.87	.89	.35	.43	.40	.40	.31	.34	.27	.18	.26	.14	.23	.70	.45	.49	.41	.26	.62				
INV	.51	.51	.42	.37	.41	.46	.70	.54	.69	.73	.68	.39	.70	.44	.58	.41	.45	.59	.70	.86	.67	.67	.93	.50			
ICG	.76	.77	.58	.48	.54	.62	.71	.55	.73	.52	.49	.38	.38	.30	.35	.28	.31	.88	.71	.76	.55	.40	.94	.66	.81		
ICC	.39	.38	.41	.34	.36	.36	.52	.43	.43	.51	.66	.62	.74	.68	.66	.64	.67	.44	.54	.57	.81	.84	.76	.41	.81	.57	
Media	9.8	9.9	10.0	10.3	10.1	10.1	10.1	9.8	9.9	10.0	9.9	10.1	10.0	10.1	10.4	9.9	10.2	19.7	19.9	19.9	20.0	20.1	59.6	20.2	49.9	39.6	40.1
Dt	2.8	3.0	3.1	3.1	2.8	3.1	2.7	3.0	3.2	3.1	2.9	3.1	2.9	3.0	3.1	2.9	3.0	5.1	5.0	5.2	4.7	5.0	12.0	5.2	10.4	8.7	8.0

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Tabla A.6. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 5:0-5:5

Prueba/puntuación de procesamiento/ puntuación compuesta	I	S	V	CO	D	N	C	RO	M	CON	R	L	BA	CA	CF	CAA	CAE	ICV	IVE	IRF	IMT	IVP	CIT	IAV	INV	ICG	ICC
Información																		.63					.68			.72	
Semejanzas	.63																	.63					.60			.60	
Vocabulario	.65	.59																									
Comprensión	.47	.51	.52																								
Dibujos	.61	.39	.59	.47																				.59			
Nombres	.66	.52	.56	.39	.59																			.59			
Cubos	.54	.39	.53	.41	.58	.50													.57				.64		.66	.58	
Rompecabezas	.31	.35	.42	.26	.42	.37	.57												.57								
Matrices	.57	.50	.49	.41	.41	.53	.52	.39												.57			.65		.64	.64	
Conceptos	.52	.48	.54	.42	.45	.44	.54	.33	.57											.57					.59		
Reconocimiento	.39	.44	.39	.30	.40	.38	.51	.38	.45	.37											.35		.62		.61		.59
Localización	.42	.23	.42	.29	.43	.33	.40	.34	.40	.34	.35										.35						.39
Búsqueda de animales	.46	.38	.48	.41	.44	.40	.48	.53	.47	.38	.59	.38										.52	.62		.61		.66
Cancelación	.32	.26	.32	.25	.28	.29	.45	.42	.45	.27	.46	.25	.52									.52					.51
Clave de figuras	.30	.29	.36	.24	.24	.31	.41	.30	.45	.39	.38	.38	.52	.63													
Cancelación aleatoria	.34	.23	.32	.24	.24	.30	.43	.44	.45	.26	.44	.21	.48	.93	.55												
Cancelación estructurada	.29	.26	.28	.25	.33	.28	.43	.35	.37	.25	.42	.27	.52	.93	.63	.76											
ICV	.90	.90	.69	.54	.55	.65	.52	.38	.59	.55	.46	.37	.47	.32	.33	.31	.30										
IVE	.48	.42	.54	.38	.56	.49	.88	.89	.51	.49	.50	.41	.57	.49	.40	.49	.44	.51									
IRF	.61	.55	.58	.47	.48	.55	.60	.41	.88	.89	.46	.41	.48	.40	.47	.40	.35	.64	.57								
IMT	.49	.40	.50	.36	.51	.43	.55	.43	.52	.43	.80	.84	.58	.43	.46	.39	.42	.50	.56	.53							
IVP	.44	.36	.45	.37	.41	.40	.53	.54	.52	.37	.59	.36	.86	.89	.66	.82	.85	.44	.60	.50	.57						
CIT	.79	.73	.69	.55	.62	.66	.76	.56	.77	.63	.75	.48	.74	.54	.52	.53	.51	.84	.75	.79	.75	.73					
IAV	.71	.51	.65	.48	.89	.90	.60	.44	.53	.50	.43	.43	.47	.32	.31	.30	.34	.68	.59	.58	.52	.45	.72				
INV	.64	.57	.63	.51	.59	.59	.79	.57	.78	.75	.76	.48	.75	.56	.56	.54	.51	.67	.77	.86	.75	.74	.95	.66			
ICG	.85	.78	.71	.56	.62	.69	.76	.52	.81	.66	.56	.46	.56	.46	.45	.45	.42	.91	.72	.82	.62	.58	.95	.73	.87		
ICC	.53	.43	.53	.41	.52	.47	.61	.55	.59	.45	.79	.68	.81	.74	.63	.68	.71	.53	.65	.58	.89	.88	.83	.55	.84	.67	
Media	10.0	10.4	10.3	10.2	9.7	10.2	9.9	10.1	10.1	9.9	9.6	10.1	10.0	9.9	10.2	10.1	9.9	20.3	20.0	19.9	19.7	19.9	59.9	19.9	49.5	40.3	39.6
Dt	2.9	2.9	3.0	3.2	3.0	3.2	2.9	2.9	3.0	3.1	3.1	3.4	2.8	3.1	3.1	3.3	3.2	5.3	5.2	5.4	5.4	5.2	13.4	5.5	11.4	9.4	9.3

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Tabla A.7. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 6:0-8:11

Prueba/puntuación de procesamiento/ puntuación compuesta	I	S	V	CO	D	N	C	RO	M	CON	R	L	BA	CA	CF	CAA	CAE	ICV	IVE	IRF	IMT	IVP	CIT	IAV	INV	ICG	ICC
Información																		.60					.60			.62	
Semejanzas	.60																	.60					.64			.61	
Vocabulario	.59	.61																									
Comprensión	.49	.52	.60																								
Dibujos	.41	.35	.32	.36																				.41			
Nombres	.41	.33	.53	.32	.41																			.41			
Cubos	.46	.39	.40	.29	.35	.32													.51				.60		.55	.53	
Rompecabezas	.30	.28	.40	.43	.40	.27	.51												.51								
Matrices	.42	.47	.48	.40	.39	.23	.45	.50												.62			.56		.62	.55	
Conceptos	.40	.37	.40	.47	.38	.17	.29	.38	.62											.62					.56		
Reconocimiento	.31	.46	.44	.38	.20	.24	.48	.30	.44	.37											.32		.54		.54		.45
Localización	.27	.26	.25	.37	.30	.13	.34	.30	.35	.31	.32										.32						.44
Búsqueda de animales	.43	.42	.40	.43	.36	.32	.43	.41	.30	.38	.34	.40										.54	.52		.48		.57
Cancelación	.26	.38	.26	.35	.24	.19	.24	.26	.33	.39	.39	.31	.54									.54					.55
Clave de figuras	.36	.35	.43	.49	.25	.20	.35	.37	.30	.34	.33	.44	.73	.52													
Cancelación aleatoria	.25	.42	.30	.39	.29	.25	.29	.35	.38	.43	.37	.31	.53	.92	.48												
Cancelación estructurada	*.26*	*.28*	*.21*	*.28*	*.18*	*.12*	*.19*	*.16*	*.27*	*.33*	*.36*	*.29*	*.50*	*.93*	*.49*	*.75*											
ICV	.90	.89	.67	.57	.42	.42	.47	.32	.50	.43	.42	.30	.47	.36	.39	.37	.30										
IVE	.44	.38	.46	.41	.44	.34	.87	.87	.55	.38	.45	.37	.48	.29	.42	.37	.20	.46									
IRF	.45	.47	.49	.49	.43	.23	.41	.49	.90	.90	.45	.37	.38	.40	.36	.45	.34	.51	.52								
IMT	.36	.45	.43	.46	.30	.23	.51	.37	.49	.42	.85	.77	.45	.44	.47	.42	.40	.45	.50	.51							
IVP	.40	.45	.38	.45	.34	.29	.39	.38	.36	.44	.42	.41	.88	.87	.71	.82	.81	.48	.44	.45	.51						
CIT	.74	.77	.67	.58	.48	.43	.74	.53	.71	.56	.71	.45	.68	.50	.56	.52	.43	.84	.73	.70	.72	.67					
IAV	.49	.40	.50	.40	.85	.83	.40	.41	.37	.33	.26	.26	.41	.26	.27	.32	.18	.50	.46	.39	.32	.38	.54				
INV	.55	.58	.58	.55	.46	.35	.72	.58	.77	.73	.73	.47	.68	.52	.57	.55	.46	.63	.74	.83	.75	.69	.93	.49			
ICG	.80	.79	.68	.55	.49	.42	.74	.51	.76	.54	.54	.39	.51	.39	.44	.43	.33	.89	.72	.72	.58	.52	.95	.54	.85		
ICC	.43	.52	.47	.52	.37	.30	.51	.43	.49	.50	.72	.67	.78	.77	.69	.73	.71	.53	.54	.55	.85	.88	.80	.40	.83	.63	
Media	9.7	9.8	9.9	9.6	9.8	9.5	9.7	10.2	9.9	9.5	9.9	9.6	10.2	10.0	9.8	9.7	10.0	19.5	19.9	19.5	19.5	20.2	59.3	19.4	49.3	39.2	39.7
Dt	3.1	3.0	3.1	2.7	3.0	2.8	2.9	3.0	3.0	3.1	3.3	2.7	3.1	3.0	2.9	3.1	3.1	5.4	5.2	5.5	4.9	5.4	13.3	4.8	11.3	9.3	8.9

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Tabla A.8. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 6:0-6:11

Prueba/puntuación de procesamiento/ puntuación compuesta	I	S	V	CO	D	N	C	RO	M	CON	R	L	BA	CA	CF	CAA	CAE	ICV	IVE	IRF	IMT	IVP	CIT	IAV	INV	ICG	ICC
Información																		.39					.50			.45	
Semejanzas	.39																	.39					.45			.46	
Vocabulario	.32	.44																									
Comprensión	.24	.26	.41																					.54			
Dibujos	.36	.26	.35	.38																				.54			
Nombres	.55	.41	.50	.33	.54														.50				.48		.47	.47	
Cubos	.31	.33	.11	.15	.32	.28													.50								
Rompecabezas	.27	.38	.19	.18	.30	.30	.50													.37			.45		.43	.49	
Matrices	.32	.35	.17	.02	.23	.26	.43	.18												.37					.46		
Conceptos	.30	.31	.22	.34	.31	.33	.36	.31	.37												.39		.42		.35		.51
Reconocimiento	.35	.31	.26	.31	.31	.27	.18	.22	.22	.24											.39						.38
Localización	.39	.36	.34	.24	.38	.38	.28	.34	.26	.21	.39											.27	.34		.38		.39
Búsqueda de animales	.24	.12	.18	.20	.43	.13	.30	.18	.14	.26	.34	.24										.27					.38
Cancelación	.10	-.02	.06	.06	.17	.09	.14	.06	.19	.13	.35	.21	.27														
Clave de figuras	.17	.25	.26	.28	.27	.21	.30	.24	.16	.29	.35	.25	.45	.31													
Cancelación aleatoria	.13	.00	.01	-.05	.15	.12	.14	.06	.18	.14	.29	.17	.20	.91	.21												
Cancelación estructurada	.06	-.04	.08	.17	.18	.04	.14	.09	.17	.10	.32	.22	.27	.89	.32	.65											
ICV	.80	.87	.46	.30	.37	.57	.38	.40	.40	.37	.39	.45	.21	.04	.25	.07	.01										
IVE	.34	.41	.17	.19	.36	.34	.88	.86	.36	.38	.23	.35	.27	.12	.32	.11	.13	.45									
IRF	.38	.40	.23	.20	.32	.35	.48	.29	.86	.80	.28	.29	.24	.20	.27	.20	.16	.47	.44								
IMT	.45	.40	.36	.33	.41	.39	.27	.33	.29	.27	.84	.82	.35	.34	.36	.28	.32	.50	.35	.34							
IVP	.22	.07	.15	.17	.38	.14	.28	.15	.21	.25	.43	.28	.81	.79	.48	.68	.72	.16	.25	.27	.43						
CIT	.67	.66	.39	.31	.50	.49	.68	.46	.64	.48	.62	.50	.56	.27	.44	.24	.24	.79	.66	.68	.68	.52					
IAV	.52	.38	.48	.40	.88	.87	.34	.34	.28	.36	.33	.43	.33	.15	.27	.15	.13	.53	.40	.38	.45	.30	.56				
INV	.47	.43	.28	.31	.49	.39	.71	.43	.66	.65	.61	.42	.63	.34	.47	.29	.31	.54	.66	.79	.63	.61	.92	.50			
ICG	.68	.73	.36	.23	.40	.52	.74	.47	.73	.47	.36	.45	.28	.14	.31	.15	.11	.85	.70	.73	.48	.27	.92	.52	.79		
ICC	.39	.28	.31	.30	.47	.31	.32	.28	.29	.31	.76	.66	.68	.66	.49	.56	.61	.40	.35	.36	.85	.84	.71	.45	.73	.45	
Media	10.0	9.7	9.8	9.7	10.5	9.9	10.4	10.0	10.0	10.7	10.0	10.3	9.7	10.1	9.8	10.1	9.9	19.7	20.4	20.7	20.3	19.8	59.8	20.4	50.7	40.1	40.1
Dt	2.5	3.0	3.0	2.7	3.0	2.8	3.1	2.9	2.8	2.4	2.8	2.7	2.8	2.7	2.8	2.8	2.7	4.7	5.2	4.3	4.6	4.4	10.9	5.1	9.1	8.2	7.6

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Tabla A.9. Intercorrelaciones de las puntuaciones de las pruebas, de las puntuaciones de procesamiento y de las puntuaciones compuestas para las edades 7:0-7:7

Prueba/puntuación de procesamiento/puntuación compuesta	I	S	V	CO	D	N	C	RO	M	CON	R	L	BA	CA	CF	CAA	CAE	ICV	IVE	IRF	IMT	IVP	CIT	IAV	INV	ICG	ICC
Información																		.42					.53			.51	
Semejanzas	.42																	.42					.54			.54	
Vocabulario	.37	.51																									
Comprensión	.33	.50	.55																								
Dibujos	.56	.48	.45	.43																				.58			
Nombres	.59	.44	.42	.44	.58																			.58			
Cubos	.32	.40	.46	.37	.41	.38													.25				.55		.56	.52	
Rompecabezas	.38	.17	.37	.26	.31	.36	.25												.25								
Matrices	.50	.48	.48	.51	.49	.49	.55	.44												.54			.67		.64	.67	
Conceptos	.47	.35	.47	.45	.53	.58	.43	.36	.54											.54					.60		
Reconocimiento	.42	.31	.34	.32	.38	.32	.26	.49	.41	.42											.44		.49		.47		.46
Localización	.37	.14	.18	.25	.26	.25	.27	.35	.35	.32	.44										.44						.56
Búsqueda de animales	.26	.32	.37	.23	.36	.35	.42	.27	.38	.38	.36	.41										.57	.47		.51		.57
Cancelación	.33	.21	.19	.30	.28	.32	.27	.20	.33	.23	.31	.45	.57									.57					.57
Clave de figuras	.33	.28	.47	.35	.36	.37	.41	.42	.37	.34	.56	.49	.62	.47													
Cancelación aleatoria	.32	.18	.14	.27	.25	.34	.21	.14	.32	.25	.28	.41	.48	.92	.37												
Cancelación estructurada	.34	.24	.22	.29	.27	.29	.32	.24	.33	.19	.35	.46	.59	.93	.53	.74											
ICV	.87	.81	.51	.48	.62	.62	.42	.34	.58	.49	.44	.31	.34	.33	.36	.31	.35										
IVE	.44	.37	.53	.40	.46	.47	.83	.75	.63	.50	.46	.39	.44	.30	.52	.23	.36	.48									
IRF	.55	.48	.54	.55	.58	.60	.57	.46	.91	.84	.47	.38	.43	.32	.41	.33	.31	.62	.65								
IMT	.47	.26	.31	.33	.38	.34	.31	.50	.45	.44	.84	.85	.46	.45	.62	.41	.47	.44	.50	.50							
IVP	.33	.30	.31	.30	.36	.38	.39	.27	.40	.34	.38	.49	.89	.88	.61	.78	.85	.38	.42	.42	.51						
CIT	.71	.69	.60	.54	.64	.62	.70	.48	.79	.62	.65	.47	.64	.48	.61	.43	.52	.83	.76	.82	.66	.64					
IAV	.64	.52	.49	.49	.89	.89	.45	.38	.55	.63	.40	.29	.40	.34	.41	.33	.31	.70	.52	.66	.40	.42	.71				
INV	.54	.51	.58	.51	.59	.58	.74	.50	.80	.73	.67	.49	.70	.47	.63	.43	.49	.62	.79	.87	.68	.66	.94	.66			
ICG	.75	.73	.59	.55	.64	.63	.74	.41	.83	.59	.46	.37	.45	.38	.46	.34	.41	.88	.74	.82	.49	.47	.95	.71	.85		
ICC	.46	.32	.36	.36	.42	.41	.41	.44	.48	.45	.70	.77	.78	.77	.71	.69	.77	.47	.53	.53	.87	.87	.75	.47	.77	.55	
Media	10.5	10.2	10.4	10.2	10.4	10.3	10.3	10.2	10.3	10.4	10.1	10.0	10.0	10.1	10.1	10.0	9.9	20.8	20.5	20.7	20.2	20.1	61.6	20.7	51.2	41.4	40.2
Dt	3.7	3.0	2.8	3.2	3.1	3.1	3.3	2.9	3.3	2.6	3.1	3.2	3.2	3.1	3.1	3.0	3.1	5.7	4.9	5.2	5.4	5.5	13.8	5.5	11.3	10.2	9.5

Nota: Los coeficientes sin corregir aparecen por debajo de la diagonal y los coeficientes corregidos aparecen por encima de la diagonal, en el área sombreada.

Uso del IAV y del INV

Introducción

El IAV es un índice secundario disponible para las dos franjas de edad de la WPPSI-IV y se basa en las pruebas *Dibujos* y *Nombres*. Conceptualmente, el IAV proporciona una estimación de la adquisición de vocabulario a partir de las destrezas del vocabulario receptivo y del expresivo. Este índice figuraba en la WPPSI-III con la denominación *Lenguaje general*, pero se ha renombrado como *Índice de adquisición de vocabulario* a fin de reflejar de forma más precisa las aptitudes que mide.

El INV también es un índice secundario disponible para las dos franjas de edad de la WPPSI-IV. Para las edades de 2:6 a 3:11 años, se basa en cuatro puntuaciones principales (*Cubos, Rompecabezas, Reconocimiento* y *Localización*) que no requieren respuestas expresivas; y para las edades de 4:0 a 7:7 años, se basa en cinco puntuaciones principales (*Cubos, Matrices, Conceptos, Reconocimiento* y *Búsqueda de animales*) que tampoco requieren respuestas expresivas. Conceptualmente, el INV ofrece una estimación de la aptitud general en las casos clínicos que presentan problemas del lenguaje expresivo (p. ej., niños con trastornos del lenguaje o trastornos del espectro autista). El INV es un índice creado para la WPPSI-IV.

En este anexo se presentan los posibles usos de estos índices secundarios, las tablas para convertir las sumas de las puntuaciones escalares en puntuaciones típicas del IAV y el INV, y las directrices básicas para su interpretación.

Aplicaciones del IAV y del INV

El IAV se diseñó principalmente como una medida de la aptitud de adquisición de vocabulario, asimismo puede ser útil como una medida de la aptitud verbal, para evaluar a niños con posibles retrasos del lenguaje. Tanto *Dibujos* como *Nombres* se basan más en el reconocimiento de vocabulario que en el conocimiento de los significados de las palabras o en las aptitudes morfológicas. Además, *Dibujos* no requiere respuestas verbales, y *Nombres* requiere solo respuestas verbales limitadas. El uso del IAV con niños con posibles retrasos del lenguaje está respaldado teóricamente por las saturaciones consistentes que estas pruebas tienen en el factor Comprensión verbal para todas las edades. Para las edades de 4:0 a 7:7 años, la validez de constructo del IAV como una medida más sencilla y más centrada en las aptitudes verbales se demuestra, además, por la presencia de dos subfactores dentro del ICV, estando uno de ellos formado por las pruebas *Dibujos* y *Nombres*, y el otro por el resto de pruebas, que tienen mayores requisitos expresivos y conceptuales (véase la figura 6.2 de este manual). Una puntuación baja en el IAV, junto con datos confirmatorios procedentes de otras fuentes (p. ej., resultados de tests de desarrollo del lenguaje, opiniones de los cuidadores o profesores), puede indicar que es necesario derivar al niño a un logopeda. No obstante, la adquisición de vocabulario es solo un aspecto concreto del funcionamiento del lenguaje; por ello, aunque el rendimiento del niño en el IAV se sitúe en el rango promedio, no puede descartarse la posibilidad de que exista un retraso o trastorno del lenguaje.

El INV ofrece una estimación de la aptitud general de los niños con retrasos del lenguaje graves y con trastornos clínicos asociados a problemas del lenguaje (p. ej., niños sordos o con dificultades

de audición o niños a los que se le ha diagnosticado trastorno del espectro autista), así como de los niños que están aprendiendo español. Dado que el INV se calcula a partir de las puntuaciones de pruebas de dos o más dominios cognitivos, es una medida más global de la aptitud intelectual general que el IVE o el IRF en aquellos casos en que, al existir problemas del lenguaje, el CI total o el ICG pueden ofrecer una estimación inadecuada. Según los datos presentados en el capítulo 5 de este manual, la fiabilidad del INV es similar a la del CI total. Los estudios descritos en el capítulo 6 de este manual demuestran la validez del INV mediante las correlaciones con otras puntuaciones compuestas de la WPPSI-IV (p. ej., el CI total) o con puntuaciones de otras medidas de la aptitud intelectual (p. ej., el Índice de aptitud Naglieri del NNAT-2). En dicho capítulo también se presentan diversos estudios con grupos especiales, representativos de trastornos o situaciones en los que el uso del INV es relevante (p. ej., discapacidad intelectual moderada, trastorno del lenguaje expresivo, trastorno autista), que aportan más datos sobre la validez del INV. Finalmente, el INV ofrece una precisión aceptable para clasificar a los niños con discapacidad intelectual o retraso del desarrollo cognitivo, y parece ser adecuado también para otros tipos de trastornos (p. ej., trastornos del lenguaje, trastornos del espectro autista). En el capítulo 6 se exponen detalladamente las distintas evidencias de validez del INV.

Es importante señalar que el INV debe utilizarse con cierta cautela puesto que su fin no es ser una vía rápida para obtener una estimación del factor *g* ni un CI total abreviado para los niños que no presentan problemas del lenguaje. Las aptitudes de comprensión verbal son componentes críticos de la aptitud intelectual; siempre que sea posible, es mejor usar el CI total como estimación de la aptitud intelectual general. Si se considera que es más apropiado utilizar el INV por cuestiones relacionadas con la competencia en el español, debería obtenerse más información sobre las aptitudes verbales del niño en su(s) otra(s) lengua(s) habitual(es). Si se elige el INV porque el niño puede presentar un retraso o trastorno del lenguaje, también se debe calcular e interpretar el IAV, si es posible, para ofrecer más información al profesional que reciba el informe de evaluación, que, en estos casos, debería ser un logopeda.

Obtención del IAV y del INV

El procedimiento para calcular el IAV y el INV se describe en el capítulo 2 del *Manual de aplicación y corrección*. La obtención de estos índices es opcional y debe basarse en el motivo de consulta o en necesidades específicas. La tabla B.1 de este manual presenta las puntuaciones del IAV equivalentes a las sumas de las puntuaciones escalares para las edades de 2:6 a 7:7 años y las tablas B.2 y B.3 presentan las puntuaciones del INV equivalentes a las sumas de las puntuaciones escalares para las edades de 2:6 a 7:7 años y de 4:0 a 7:7 años, respectivamente.

Tabla B.1. Conversión suma de puntuaciones escalares a IAV para las edades de 2:6 a 7:7 años (datos españoles)

Suma de puntuaciones escalares	IAV	Percentil	Intervalo de confianza	
			90%	95%
2	45	<0.1	44-60	43-61
3	49	<0.1	48-63	46-65
4	52	0.1	50-66	49-67
5	55	0.1	53-69	52-70
6	58	0.3	56-71	54-73
7	61	0.5	58-74	57-75
8	64	1	61-76	59-78
9	67	1	64-79	62-81
10	70	2	66-82	65-83
11	73	4	69-84	67-86
12	76	5	71-87	70-88
13	79	8	74-89	73-91
14	82	12	77-92	75-94
15	85	16	79-95	78-96
16	88	21	82-97	80-99
17	91	27	84-100	83-101
18	94	34	87-103	86-104
19	97	42	90-105	88-107
20	100	50	92-108	91-109
21	103	58	95-110	93-112
22	106	66	97-113	96-114
23	109	73	100-116	99-117
24	112	79	103-118	101-120
25	115	84	105-121	104-122
26	118	88	108-123	106-125
27	120	91	110-125	108-127
28	122	93	111-127	110-128
29	125	95	114-129	113-131
30	128	97	117-132	115-134
31	131	98	119-135	118-136
32	135	99	123-138	121-140
33	138	99	125-141	124-142
34	141	99.7	128-143	126-145
35	144	99.8	131-146	129-148
36	147	99.9	133-149	132-150
37	151	>99.9	137-152	135-154
38	155	>99.9	140-156	139-157

Tabla B.2. Conversión suma de puntuaciones escalares a INV para las edades de 2:6 a 3:11 años (datos españoles)

Suma de puntuaciones escalares	INV	Percentil	Intervalo de confianza 90%	Intervalo de confianza 95%	Suma de puntuaciones escalares	INV	Percentil	Intervalo de confianza 90%	Intervalo de confianza 95%
4	40	<0.1	38-51	37-52	41	102	55	95-108	94-109
5	42	<0.1	40-53	39-54	42	104	61	97-110	96-111
6	43	<0.1	41-54	40-55	43	105	63	98-111	97-112
7	45	<0.1	43-56	42-57	44	107	68	100-113	99-114
8	47	<0.1	45-58	44-59	45	109	73	102-115	101-116
9	49	<0.1	47-60	45-61	46	110	75	103-116	102-117
10	51	0.1	48-61	47-63	47	112	79	105-117	103-119
11	52	0.1	49-62	48-63	48	114	82	106-119	105-121
12	54	0.1	51-64	50-65	49	116	86	108-121	107-122
13	56	0.2	53-66	52-67	50	118	88	110-123	109-124
14	58	0.3	55-68	54-69	51	120	91	112-125	111-126
15	59	0.3	56-69	55-70	52	122	93	114-127	113-128
16	61	0.5	58-71	56-72	53	123	94	115-128	114-129
17	62	1	59-71	57-73	54	125	95	117-129	115-131
18	63	1	60-72	58-74	55	127	96	118-131	117-132
19	65	1	61-74	60-75	56	128	97	119-132	118-133
20	66	1	62-75	61-76	57	130	98	121-134	120-135
21	67	1	63-76	62-77	58	131	98	122-135	121-136
22	69	2	65-78	64-79	59	133	99	124-137	123-138
23	70	2	66-79	65-80	60	134	99	125-138	124-139
24	72	3	68-81	67-82	61	136	99	127-140	125-141
25	73	4	69-82	68-83	62	137	99	128-140	126-142
26	74	4	70-83	68-84	63	139	99.5	129-142	128-144
27	76	5	71-84	70-86	64	141	99.7	131-144	130-145
28	78	7	73-86	72-87	65	142	99.7	132-145	131-146
29	79	8	74-87	73-88	66	143	99.8	133-146	132-147
30	81	10	76-89	75-90	67	145	99.9	135-148	134-149
31	83	13	78-91	77-92	68	146	99.9	136-149	135-150
32	85	16	80-93	79-94	69	148	99.9	138-151	137-152
33	87	19	82-94	80-96	70	149	99.9	139-152	137-153
34	89	23	83-96	82-98	71	151	>99.9	140-153	139-155
35	91	27	85-98	84-99	72	152	>99.9	141-154	140-155
36	93	32	87-100	86-101	73	154	>99.9	143-156	142-157
37	94	34	88-101	87-102	74	156	>99.9	145-158	144-159
38	96	39	90-103	89-104	75	158	>99.9	147-160	146-161
39	98	45	92-105	91-106	76	160	>99.9	149-162	148-163
40	100	50	94-106	92-108					

Tabla B.3. Conversión suma de puntuaciones escalares a INV para las edades de 4:0 a 7:7 años (datos españoles)

Suma de puntuaciones escalares	INV	Percentil	Intervalo de confianza 90%	Intervalo de confianza 95%	Suma de puntuaciones escalares	INV	Percentil	Intervalo de confianza 90%	Intervalo de confianza 95%
5	40	<0.1	38-51	37-52	51	101	53	94-107	93-109
6	41	<0.1	39-52	38-53	52	103	58	96-109	95-110
7	43	<0.1	41-54	40-55	53	104	63	97-110	96-111
8	44	<0.1	42-55	41-56	54	106	66	99-112	98-113
9	46	<0.1	44-57	43-58	55	107	70	100-113	99-114
10	47	<0.1	45-58	44-59	56	109	73	102-115	101-116
11	49	<0.1	47-60	45-61	57	110	75	103-116	102-117
12	50	0.1	48-60	46-62	58	112	79	105-117	103-119
13	52	0.1	49-62	48-63	59	113	81	106-118	104-120
14	54	0.1	51-64	50-65	60	115	84	107-120	106-121
15	55	0.1	52-65	51-66	61	116	86	108-121	107-122
16	56	0.2	53-66	52-67	62	118	87	110-123	109-124
17	57	0.3	54-67	53-68	63	119	90	111-124	110-125
18	58	0.4	55-68	54-69	64	121	91	113-126	112-127
19	60	0.5	57-70	56-71	65	122	93	114-127	113-128
20	61	1	58-71	56-72	66	124	95	116-129	114-130
21	62	1	59-71	57-73	67	125	96	117-129	115-131
22	63	1	60-72	58-74	68	126	96	117-130	116-132
23	65	1	61-74	60-75	69	128	97	119-132	118-133
24	66	1	62-75	61-76	70	129	98	120-133	119-134
25	67	2	63-76	62-77	71	131	98	122-135	121-136
26	68	2	64-77	63-78	72	132	99	123-136	122-137
27	69	2	65-78	64-79	73	133	99	124-137	123-138
28	71	3	67-80	66-81	74	135	99	126-139	125-140
29	72	3	68-81	67-82	75	136	99	127-140	125-141
30	73	4	69-82	68-83	76	138	99.5	129-141	127-143
31	74	4	70-83	68-84	77	139	99.6	129-142	128-144
32	75	5	71-83	69-85	78	140	99.7	130-143	129-144
33	76	6	71-84	70-86	79	142	99.8	132-145	131-146
34	77	7	72-85	71-86	80	143	99.8	133-146	132-147
35	78	8	73-86	72-87	81	144	99.9	134-147	133-148
36	79	9	74-87	73-88	82	145	99.9	135-148	134-149
37	81	10	76-89	75-90	83	146	99.9	136-149	135-150
38	82	12	77-90	76-91	84	147	99.9	137-150	136-151
39	84	14	79-92	78-93	85	148	99.9	138-151	137-152
40	85	16	80-93	79-94	86	149	>99.9	139-152	137-153
41	87	18	82-94	80-96	87	150	>99.9	140-152	138-154
42	88	19	83-95	81-97	88	151	>99.9	140-153	139-155
43	89	23	83-96	82-98	89	152	>99.9	141-154	140-155
44	90	25	84-97	83-98	90	153	>99.9	142-155	141-156
45	92	30	86-99	85-100	91	154	>99.9	143-156	142-157
46	93	32	87-100	86-101	92	155	>99.9	144-157	143-158
47	95	37	89-102	88-103	93	157	>99.9	146-159	145-160
48	96	39	90-103	89-104	94	158	>99.9	147-160	146-161
49	98	45	92-105	91-106	95	160	>99.9	149-162	148-163
50	100	50	94-106	92-108					

Interpretación básica del IAV y del INV

La información que sigue es una guía básica para interpretar el IAV y el INV. El uso de estos índices secundarios es opcional, basado en el motivo de la consulta o en necesidades específicas, y debe realizarse como parte del análisis secundario del perfil, tal y como se describe en el capítulo 7 de este manual. Los aspectos comunes de la interpretación del IAV y del INV, como los percentiles, el error típico de medida y los intervalos de confianza, y la clasificación cualitativa, se exponen en el capítulo 7 de este manual.

Interpretación del IAV

Una puntuación baja en el IAV puede indicar la presencia de problemas del lenguaje expresivo relacionados con una afección clínica (p. ej., trastorno del lenguaje expresivo). Nunca debería utilizarse este índice para diagnosticar un trastorno del lenguaje, puesto que solo cubre un área restringida del desarrollo del lenguaje. Ante la sospecha de un trastorno del lenguaje, debería realizarse una evaluación completa o derivar al niño a un logopeda.

Es posible que los niños que están aprendiendo español como segunda lengua obtengan puntuaciones más bajas en el IAV que los niños cuya única lengua es el español. Por ello, una puntuación baja en el IAV puede indicar a los profesionales que el niño se encuentra en esta situación; en tal caso, debe tenerse en cuenta que el IAV no pretende medir la competencia en el español ni de los niños bilingües ni de los que lo están aprendiendo. Los ítems de baja dificultad de las pruebas *Dibujos* y *Nombres* se corresponden con palabras conocidas por la mayoría de los niños. Si el niño entiende la tarea y habla español, debería poder responder correctamente a estos ítems.

La evaluación de los niños que están aprendiendo una segunda lengua es un proceso complejo. Es necesario conocer la competencia de un niño en su lengua materna y en otras lenguas, así como estar familiarizado con los distintos modelos utilizados para adaptar dicha evaluación a la situación del niño (Ortiz et al., 2012). Por ello, la identificación del nivel de conocimiento y de competencia de un niño en una segunda lengua, como un aspecto importante, así como su evaluación, debería llevarlas a cabo un profesional con experiencia en este ámbito concreto.

El rendimiento en el IAV también puede ser menor en los niños de niveles sociales más bajos. Por ejemplo, los tipos de conversación que presencian los niños cuyos padres tienen niveles educativos más altos son distintos a los que presencian los niños cuyos padres tienen niveles educativos más bajos. Esta diferencia parece afectar a la adquisición de vocabulario de los niños (Hart y Risely, 2003).

Una puntuación baja en el IAV también puede indicar que es conveniente realizar evaluaciones adicionales con otras medidas, puesto que un nivel bajo de vocabulario puede conllevar problemas de comprensión lectora (Kendeou et al., 2012); es decir, algunos niños pueden decodificar las palabras pero no entender su significado porque tienen un nivel de vocabulario insuficiente. Las actividades de enriquecimiento de vocabulario son útiles como parte de un plan de tratamiento global de los problemas de comprensión lectora.

El IAV debe evaluarse junto con las puntuaciones de las pruebas que lo conforman; por ello, para interpretar este índice, es relevante comparar las puntuaciones de *Dibujos* y *Nombres*. Esta comparación permite determinar si existen diferencias significativas entre la adquisición de vocabulario receptivo y la del vocabulario expresivo, así como con qué frecuencia se observan estas diferencias en la muestra de tipificación. Una puntuación en *Dibujos* significativamente más alta que la obtenida en *Nombres* puede sugerir que la adquisición de vocabulario receptivo está mucho más desarrollada que la adquisición del vocabulario expresivo, o que existe un problema con la recuperación de las palabras. En cambio, una puntuación en *Nombres* significativamente más alta que la obtenida en *Dibujos* puede sugerir que el desarrollo del lenguaje expresivo está más avanzado que el desarrollo del lenguaje receptivo respecto a los niños de la misma edad. Dado que algunos de los estímulos de *Dibujos* son más complejos que los estímulos de *Nombres*, podrían descartarse los problemas de agudeza visual o de atención. Asimismo, es más frecuente que las respuestas impulsivas se den en *Dibujos* puesto que es una prueba con formato de opción múltiple. Dado que estas pruebas cubren solo determinadas áreas del desarrollo del lenguaje y *Nombres* solo requiere una

única palabra como respuesta, la interpretación de las diferencias entre *Dibujos* y *Nombres* como indicadoras de posibles problemas del lenguaje es limitada y debería ratificarse con otros datos. En la página de análisis secundario del cuadernillo de anotación se incluye una tabla que permite realizar la comparación entre las puntuaciones de *Dibujos* y *Nombres*. En el capítulo 7 se ofrece más información sobre la comparación de estas puntuaciones.

Interpretación del INV

El INV ofrece una estimación global de la aptitud intelectual, sin incluir el dominio de la comprensión verbal. Las pruebas que contribuyen a este índice pertenecen a las escalas Visoespacial y Memoria de trabajo para la franja de edad de 2:6 a 3:11 años, y de las escalas Visoespacial, Razonamiento fluido, Memoria de trabajo y Velocidad de procesamiento para la franja de edad de 4:0 a 7:7 años. El INV puede interpretarse como una medida del funcionamiento intelectual general que reduce al mínimo la intervención del lenguaje para permitir evaluar a niños cuyas circunstancias son especiales (p. ej., niños que están aprendiendo español) o a niños con necesidades clínicas (p. ej., niños con trastornos del espectro autista). La ausencia de requisitos expresivos también hace que este índice sea útil para medir la aptitud cognitiva general de los niños sordos o con dificultades de audición. No obstante, el INV no puede considerarse exento del componente lingüístico, puesto que las pruebas que lo conforman requieren que el niño *disponga de la competencia necesaria para comprender las consignas* (Ortiz et al., 2012). Los índices como el INV pueden utilizarse junto con los procedimientos establecidos para evaluar la inteligencia de los niños que están aprendiendo español (Ortiz et al., 2012). Realizar comparaciones entre los índices primarios (excepto del ICV) y de las pruebas que los conforman puede servir para interpretar el INV cuando la variabilidad del rendimiento del niño en las pruebas que contribuyen al INV es muy grande.

Uso del ICG y del ICC

Descripción

El ICG es un índice secundario que se basa en las pruebas de Comprensión verbal, Visoespacial y (solo para la franja de edad de 4:0 a 7:7 años) Razonamiento fluido que contribuyen al CI total. Este índice puede obtenerse para las dos franjas de edad de la WPPSI-IV. Conceptualmente, el ICG ofrece una estimación de la capacidad intelectual general con una menor intervención de la memoria de trabajo y de la velocidad de procesamiento que el CI total.

El ICC es un índice secundario que se basa en las pruebas que contribuyen al IMT y al IVP, y solo puede obtenerse para las edades de 4:0 a 7:7 años. La memoria de trabajo y la velocidad de procesamiento se relacionan porque la memoria de trabajo implica la manipulación de la información para el almacenamiento a corto plazo, y la velocidad de procesamiento se relaciona, en parte, con la velocidad cognitiva antes de que la información decaiga. Conceptualmente, el ICC ofrece una estimación de la competencia con la que se procesa la información cognitiva durante el aprendizaje, la resolución de problemas y el razonamiento de nivel superior.

El ICG y el ICC son índices nuevos en esta versión de la WPPSI-IV; dado que la WPPSI-III no incluía pruebas de memoria de trabajo, y solo una prueba de velocidad de procesamiento contribuía al CI total para las edades de 4:0 a 7:3 años, estos índices eran innecesarios. No obstante, tanto el ICG como el ICC forman parte de otras escalas de inteligencia de Wechsler.

Introducción al ICG y al ICC

El ICG se creó para la WISC-III estadounidense (Prifitera, Weiss y Saklofske,1998). Ofrecía una puntuación resumida que permitía medir la capacidad intelectual general como resultado de la suma de las puntuaciones escalares de todas las pruebas que contribuían al CI total, a excepción de *Aritmética* y *Claves*. Se recomendaba usar el ICG como una estimación de la capacidad general en casos en que la diferencia entre el CI verbal y el CI manipulativo fuese muy grande debido a puntuaciones bajas en *Aritmética* y/o *Claves* (Prifitera et al., 1998). Por otra parte, el ICG de la WAIS-III estadounidense (Tulsky, Saklofske, Wilkins y Weiss, 2001) se obtenía, con un objetivo similar, a partir de las pruebas que contribuían al CI total, con excepción de las pruebas de Memoria de trabajo y de Velocidad de procesamiento.

En Estados Unidos, la WISC-IV es la primera escala que ofrece tanto el ICG como el ICC. En España, el ICG se incluye por primera vez en la WAIS-IV. En la WISC-IV, el ICG (Raiford et al., 2005; Saklofske, Prifitera, Weiss, Rolfhus y Zhu, 2005) se obtiene a partir de las pruebas de Comprensión verbal y de Razonamiento perceptivo que contribuyen al CI total, y el ICC se obtiene a partir de las pruebas de Memoria de trabajo y de Velocidad de procesamiento que contribuyen al IMT y al IVP. Por otra parte, el ICG de la WAIS-IV se basa, con un objetivo similar al del ICG de la WISC-IV, en las tres pruebas de Comprensión verbal y las tres pruebas de Razonamiento perceptivo que contribuyen al CI total. El ICC de la WAIS-IV se basa en las dos pruebas de Memoria de trabajo y las dos pruebas de Velocidad de procesamiento que contribuyen al IMT y al IVP (Weiss, Saklofske, Coalson y Raiford, 2010). El ICC refleja la competencia y eficiencia del procesamiento cognitivo y ofrece un resumen del rendimiento en tareas de memoria

de trabajo y de velocidad de procesamiento. En relación con el IMT y el IVP, el ICC es una estimación más global de la competencia del procesamiento cognitivo.

Los datos recopilados sugieren que el ICG y el ICC pueden ser útiles cuando se interpretan en el conjunto de una evaluación completa. Las investigaciones indican que el porcentaje de niños en edad escolar con trastornos del lenguaje que obtienen puntuaciones más bajas en el ICC y en medidas del rendimiento académico que en el ICG es considerablemente más alto que el porcentaje de niños del grupo control (Weiss y Gabel, 2008). También se ha observado un patrón de puntuaciones significativamente más bajas en el ICC y el ICG en niños en edad escolar con epilepsia (Gottlieb, Zelko, Kim y Nordli, 2012) y en niños con TDAH con predominio del déficit de atención, pero no con TDAH de tipo combinado (Bremner, McTaggart, Saklofske y Janzen, 2011). Los adultos con trastornos neurocognitivos (trastornos del aprendizaje, TDAH y traumatismo craneoencefálico) mostraron diferencias más grandes entre el ICC y el ICG (siendo el ICC inferior al ICG) que las mostradas por los adultos con trastornos psicológicos (Harrison, DeLisle y Parker, 2008). Sin embargo, las diferencias entre el ICG y el ICC por sí solas no deben considerarse indicadores diagnósticos del TDAH (Devena y Watkins, 2012) ni de otros trastornos.

Aplicación del ICG y del ICC

Habitualmente, el CI total se usa para describir un aspecto global subyacente de la inteligencia general (*g*). En la práctica clínica, el CI total se utiliza para diversos objetivos. Puede utilizarse como síntesis del rendimiento en los distintos dominios de aptitud cognitiva (p. ej., comprensión verbal, visoespacial, razonamiento fluido, memoria de trabajo y velocidad de procesamiento). El CI total es una puntuación global que resume, en un único número, el rendimiento de un sujeto en varias aptitudes cognitivas. Cuando se observa una diferencia poco frecuente entre los dominios o las pruebas que contribuyen al CI total, la interpretación clínica debe tener en cuenta las diferencias entre las aptitudes.

Los niños con trastornos del neurodesarrollo relacionados con problemas de la memoria de trabajo y la velocidad de procesamiento, como los trastornos del aprendizaje, el TDAH, los trastornos del lenguaje o los trastornos del espectro autista (Archibald y Gathercole, 2006a, 2006b; Belleville et al., 2006; Passolunghi, 2006; Pickering, 2006; Roodenrys, 2006; Swanson, 2006), pueden obtener un CI total más bajo que los niños sin tales problemas, dado que las áreas afectadas están incluidas en el CI total. En tales casos, el CI total bajo puede ocultar diferencias significativas entre la aptitud cognitiva general (representada por el CI total) y otras funciones cognitivas (p. ej., rendimiento, memoria y funciones neuropsicológicas específicas). El ICG se desarrolló específicamente para ayudar a los profesionales en la identificación de los puntos fuertes y débiles basados en comparaciones entre la aptitud cognitiva general y otras funciones cognitivas del sujeto. Respecto al CI total, el ICG proporciona una estimación de la aptitud intelectual general menos sensible a la influencia de las dificultades en la memoria de trabajo y la velocidad de procesamiento porque se han eliminado las pruebas correspondientes a estas áreas. Se puede comparar el CI total y el ICG para evaluar los efectos que tiene un punto débil en la competencia cognitiva (medida por las pruebas de Memoria de trabajo y Velocidad de procesamiento) sobre el funcionamiento cognitivo general del niño. En algunos casos puede ser conveniente utilizar el ICG para compararlo con medidas del rendimiento académico o de otras funciones cognitivas. La evaluación de la significación y frecuencia de las diferencias entre el ICG y el CI total puede servir para decidir cuándo se debe usar el ICG en determinadas situaciones clínicas.

Cuando se usa el ICG, debe registrarse e interpretarse junto con el CI total y las puntuaciones de todos los índices primarios, incluidos el IMT y el IVP. El ICG no sustituye al CI total como estimación de la aptitud intelectual general. La memoria de trabajo y la velocidad de procesamiento son fundamentales para realizar una evaluación completa de la aptitud cognitiva general (Blalock y McCabe, 2011; Bunting, 2006; Dodonova y Dodonov, 2012). No incluir las medidas de estas aptitudes en la puntuación resumida reduce la amplitud de constructo y la validez predictiva de dicha puntuación (Rowe et al., 2010).

Los profesionales deben tener en cuenta que las diferencias pequeñas entre el ICG y el CI total pueden ser significativas en la WPPSI-IV. Ello es debido a que el CI total y el ICG comparten el contenido de las pruebas, a excepción de una prueba para la franja de menor edad y de dos pruebas para la franja de mayor edad. En comparación, el CI total de la WISC-IV y el de la WAIS-IV incluyen cuatro pruebas que no contribuyen al ICG de dichas escalas. Por ello, las diferencias más pequeñas entre el CI total y el ICG son más significativas en la WPPSI-IV que en la WISC-IV o la WAIS-IV. Además, respecto a la WISC-IV (Raiford et al., 2005) y la WAIS-IV (Wechsler, 2008), son necesarias diferencias más pequeñas entre el CI total y el ICG de la WPPSI-IV para obtener la significación estadística.

Dado que el CI total y el ICG de la WPPSI-IV tienen muchas pruebas comunes, la comparación del ICG con el ICC puede ofrecer más información para la interpretación de los resultados obtenidos por niños de 4:0 a 7:7 meses. Esta comparación es conceptualmente distinta a la comparación del ICG con el CI total. El ICC incluye pruebas complementarias de Memoria de trabajo y de Velocidad de procesamiento que no contribuyen al CI total, lo que proporciona una mayor amplitud de constructo de la competencia cognitiva. El ICG puede compararse con el ICC para obtener más información sobre el posible efecto que los déficits cognitivos básicos pueden tener sobre la aptitud intelectual general del niño.

Cabe destacar que el 63 % de los niños con alta capacidad intelectual que participaron en el estudio con grupos especiales estadounidenses obtuvieron diferencias en las que el ICG era superior al CI total. Estos resultados se asemejan a los observados tanto en niños en edad escolar (Raiford et al., 2005) como en adolescentes y adultos (Wechsler, 2008). Además el 79,2 % de los niños con alta capacidad intelectual que participaron en el estudio con grupos especiales obtuvieron diferencias en las que el ICG era superior al ICC. Asimismo se observaron patrones similares en los niños de la muestra de tipificación que mostraron un nivel alto de aptitud general (ICG ≥ 120), lo que indica que frecuentemente la memoria de trabajo y la velocidad de procesamiento no son las áreas de rendimiento más fuertes en los niños con una aptitud intelectual alta. En la tabla 6.17 se observa que los tamaños del efecto para las diferencias del rendimiento entre el grupo de niños con alta capacidad intelectual y el grupo control son aproximadamente iguales tanto para el ICG como para el CI total, y que, aunque ambos tamaños del efecto son grandes, el tamaño del efecto del ICC es considerablemente más pequeño que el del ICG. Por todo ello, algunos autores abogan a favor de utilizar el ICG como criterio para la asignación de niños a programas especializados en alta capacidad intelectual o similares (Rimm et al., 2008). Al evaluar a niños con alta capacidad intelectual, aunque el ICG puede aportar más información relevante, es posible que los puntos débiles normativos o intrapersonales en la competencia cognitiva no resulten obvios si no se han aplicado las pruebas de memoria de trabajo y de velocidad de procesamiento.

Los estudios estadounidenses con grupos especiales, descritos en el capítulo 6 de este manual, sugieren que las diferencias entre la capacidad general y la competencia cognitiva de los niños con trastornos clínicos pueden no ser tan marcadas como las observadas entre la WISC-IV y la WAIS-IV. Esto podría deberse a varias causas. En primer lugar, es posible que las diferencias entre la capacidad general y la competencia cognitiva de los niños pequeños con trastornos clínicos no sean aún tan grandes y que se incrementen con el tiempo. En segundo lugar, las pruebas que contribuyen al IMT y al IVP de la WPPSI-IV son diferentes a las correspondientes de la WISC-IV y la WAIS-IV. En concreto, las pruebas de la WPPSI-IV miden la memoria de trabajo visual, mientras que las de la WISC-IV y la WAIS-IV incluyen tareas de memoria de trabajo verbal. Aunque la memoria de trabajo verbal y la memoria de trabajo visual estén estrechamente relacionadas, son diferencialmente sensibles a diversos trastornos del neurodesarrollo, como el trastorno de la lectura, los trastornos del lenguaje y los trastornos generalizados del desarrollo (Archibald y Gathercole, 2006a, 2006b, 2007; Borella et al., 2010; Sinzig, Morsch, Bruning, Schmidt y Lehmkuhl, 2008), y se asocian con la activación de áreas cerebrales ligeramente distintas (Jolles et al., 2011; Martin et al., 2008). Esto puede explicar parte de la sensibilidad diferencial de las pruebas de memoria de trabajo visual de la WPPSI-IV. Además, el hecho de usar un tampón en vez de un lápiz para indicar las respuestas en las pruebas de Velocidad de procesamiento reduce los requisitos de la motricidad fina en esas tareas. De este modo, en la WPPSI-IV, la variabilidad inesperada debida a déficits motores más que a la velocidad de procesamiento habituales en

varios trastornos del neurodesarrollo (p. ej., discapacidad intelectual, trastorno autista) es menor respecto a la de las pruebas de Velocidad de procesamiento de la WISC-IV y la WAIS-IV.

Por lo general, los patrones de puntos fuertes y débiles cognitivos que se esperaba observar en los grupos especiales se reflejan en los resultados obtenidos por estos en el ICG y el ICC. No obstante, estos resultados son los del grupo y no son evidentes en cada uno de los niños del grupo. Por todo ello, las comparaciones entre el CI total, el ICG y el ICC no deben considerarse indicadores diagnósticos de ningún grupo específico, sino indicadores de puntos fuertes y débiles en los conjuntos de aptitudes medidas por estas puntuaciones compuestas.

El profesional puede calcular e interpretar el ICG y el ICC en diversas situaciones clínicas, como las siguientes:

- una diferencia significativa e infrecuente entre el IMT y la MIP o el CI total;
- una diferencia significativa e infrecuente entre el IVP y la MIP o el CI total;
- una diferencia significativa e infrecuente entre el ICV y el IMT;
- una diferencia significativa e infrecuente entre el ICV y el IVP;
- una diferencia significativa e infrecuente entre el IVE y el IMT;
- una diferencia significativa e infrecuente entre el IVE y el IVP;
- una diferencia significativa e infrecuente entre el IRF y el IMT;
- una diferencia significativa e infrecuente entre el IRF y el IVP;
- una diferencia significativa e infrecuente entre el IMT y el IVP;
- una diferencia significativa e infrecuente entre las pruebas que contribuyen al IMT o al IVP;
- una diferencia significativa e infrecuente entre las pruebas de Memoria de trabajo o de Velocidad de procesamiento y la MPE-IP o la MPE-CI.

Todas las puntuaciones compuestas aquí descritas figuran en el cuadernillo de anotación. En la sección «Cumplimentar el cuadernillo de anotación» del capítulo 2 del *Manual de aplicación y corrección* se ofrece más información, y en las tablas B-1-B.8 de dicho manual figuran los datos de tipificación necesarios para comparar las puntuaciones.

Obtención del ICG y del ICC

El procedimiento para calcular el ICG y el ICC se describe en el capítulo 2 del *Manual de aplicación y corrección*. Y las tablas C.1 y C.2 de este manual presentan las puntuaciones del ICG equivalentes a las sumas de las puntuaciones escalares para las edades de 2:6 a 3:11 años y de 4:0 a 7:7 años, respectivamente. En la tabla C.3. figuran las puntuaciones del ICC equivalentes a las sumas de las puntuaciones escalares para las edades de 4:0 a 7:7 años.

Tabla C.1. Conversión suma de puntuaciones escalares a ICG para las edades de 2:6 a 3:11 años (datos españoles)

Suma de puntuaciones escalares	ICG	Percentil	Intervalo de confianza		Suma de puntuaciones escalares	ICG	Percentil	Intervalo de confianza	
			90%	95%				90%	95%
4	40	<0.1	38-51	37-52	41	102	55	95-108	94-109
5	42	<0.1	40-53	39-54	42	104	61	97-110	96-111
6	44	<0.1	42-55	41-56	43	106	66	99-112	98-113
7	46	<0.1	44-57	43-58	44	107	68	100-113	99-114
8	48	<0.1	46-59	45-60	45	109	73	102-115	101-116
9	50	<0.1	48-60	46-62	46	111	77	104-117	102-118
10	52	0.1	49-62	48-63	47	113	81	106-118	104-120
11	53	0.1	50-63	49-64	48	115	84	107-120	106-121
12	55	0.1	52-65	51-66	49	116	86	108-121	107-122
13	56	0.2	53-66	52-67	50	118	88	110-123	109-124
14	58	0.3	55-68	54-69	51	120	91	112-125	111-126
15	60	0.4	57-70	56-71	52	122	93	114-127	113-128
16	62	1	59-71	57-73	53	124	95	116-129	114-130
17	63	1	60-72	58-74	54	126	96	117-130	116-132
18	64	1	60-73	59-75	55	128	97	119-132	118-133
19	66	1	62-75	61-76	56	130	98	121-134	120-135
20	67	1	63-76	62-77	57	131	98	122-135	121-136
21	69	2	65-78	64-79	58	133	99	124-137	123-138
22	71	3	67-80	66-81	59	134	99	125-138	124-139
23	72	3	68-81	67-82	60	136	99	127-140	125-141
24	74	4	70-83	68-84	61	137	99	128-140	126-142
25	75	5	71-83	69-85	62	139	99.5	129-142	128-144
26	76	5	71-84	70-86	63	140	99.6	130-143	129-144
27	78	7	73-86	72-87	64	141	99.7	131-144	130-145
28	79	8	74-87	73-88	65	143	99.8	133-146	132-147
29	80	9	75-88	74-89	66	145	99.9	135-148	134-149
30	82	12	77-90	76-91	67	146	99.9	136-149	135-150
31	84	14	79-92	78-93	68	148	99.9	138-151	137-152
32	86	18	81-94	79-95	69	150	>99.9	140-152	138-154
33	88	21	83-95	81-97	70	151	>99.9	140-153	139-155
34	89	23	83-96	82-98	71	153	>99.9	142-155	141-156
35	91	27	85-98	84-99	72	154	>99.9	143-156	142-157
36	93	32	87-100	86-101	73	156	>99.9	145-158	144-159
37	95	37	89-102	88-103	74	157	>99.9	146-159	145-160
38	97	42	91-104	90-105	75	158	>99.9	147-160	146-161
39	99	47	93-106	91-107	76	160	>99.9	149-162	148-163
40	100	50	94-106	92-108					

Tabla C.2. Conversión suma de puntuaciones escalares a ICG para las edades de 4:0 a 7:7 años (datos españoles)

Suma de puntuaciones escalares	ICG	Percentil	*Intervalo de confianza*		Suma de puntuaciones escalares	ICG	Percentil	*Intervalo de confianza*	
			90%	95%				90%	95%
4	40	<0.1	39-52	37-53	**41**	101	53	94-108	93-109
5	42	<0.1	40-54	39-55	**42**	103	58	96-109	95-111
6	44	<0.1	42-56	41-57	**43**	104	61	97-110	96-112
7	46	<0.1	44-58	43-59	**44**	106	66	99-112	97-113
8	48	<0.1	46-59	45-61	**45**	108	70	101-114	99-115
9	50	<0.1	48-61	46-63	**46**	110	75	102-116	101-117
10	51	0.1	49-62	47-63	**47**	112	79	104-118	103-119
11	53	0.1	50-64	49-65	**48**	113	81	105-119	104-120
12	55	0.1	52-66	51-67	**49**	114	82	106-119	105-121
13	56	0.2	53-67	52-68	**50**	116	86	108-121	107-123
14	57	0.2	54-68	53-69	**51**	118	88	110-123	108-124
15	59	0.3	56-69	55-71	**52**	120	91	111-125	110-126
16	60	0.4	57-70	56-72	**53**	121	92	112-126	111-127
17	62	1	59-72	57-73	**54**	123	94	114-128	113-129
18	63	1	60-73	58-74	**55**	125	95	116-129	115-131
19	65	1	61-75	60-76	**56**	127	96	118-131	117-133
20	66	1	62-76	61-77	**57**	128	97	119-132	117-134
21	67	1	63-77	62-78	**58**	130	98	121-134	119-135
22	69	2	65-79	64-80	**59**	132	98	122-136	121-137
23	70	2	66-79	65-81	**60**	133	99	123-137	122-138
24	71	3	67-80	66-82	**61**	135	99	125-139	124-140
25	73	4	69-82	67-83	**62**	137	99	127-140	126-142
26	75	5	71-84	69-85	**63**	139	99.5	129-142	127-144
27	77	6	72-86	71-87	**64**	140	99.6	130-143	128-144
28	79	8	74-88	73-89	**65**	142	99.7	131-145	130-146
29	81	10	76-89	75-91	**66**	144	99.8	133-147	132-148
30	83	13	78-91	77-93	**67**	145	99.9	134-148	133-149
31	85	16	80-93	78-94	**68**	147	99.9	136-150	135-151
32	87	19	81-95	80-96	**69**	148	99.9	137-150	136-152
33	89	23	83-97	82-98	**70**	149	99.9	138-151	137-153
34	90	25	84-98	83-99	**71**	151	>99.9	140-153	138-154
35	91	27	85-99	84-100	**72**	153	>99.9	141-155	140-156
36	93	32	87-100	86-102	**73**	154	>99.9	142-156	141-157
37	95	37	89-102	87-103	**74**	156	>99.9	144-158	143-159
38	97	42	91-104	89-105	**75**	158	>99.9	146-160	145-161
39	99	47	92-106	91-107	**76**	160	>99.9	148-161	147-163
40	100	50	93-107	92-108					

Tabla C.3. Conversión suma de puntuaciones escalares a ICC para las edades de 4:0 a 7:7 años (datos españoles)

Suma de puntuaciones escalares	ICC	Percentil	Intervalo de confianza 90%	Intervalo de confianza 95%	Suma de puntuaciones escalares	ICC	Percentil	Intervalo de confianza 90%	Intervalo de confianza 95%
4	40	<0.1	39-53	38-54	41	102	55	95-109	93-110
5	42	<0.1	41-55	39-56	42	104	61	97-111	95-112
6	44	<0.1	43-57	41-58	43	106	66	98-112	97-114
7	45	<0.1	43-58	42-59	44	108	70	100-114	99-116
8	47	<0.1	45-59	44-61	45	110	75	102-116	101-117
9	49	<0.1	47-61	46-62	46	111	77	103-117	102-118
10	50	<0.1	48-62	47-63	47	113	81	105-119	103-120
11	51	0.1	49-63	48-64	48	115	84	106-121	105-122
12	53	0.1	51-65	49-66	49	116	86	107-121	106-123
13	54	0.1	52-66	50-67	50	118	88	109-123	108-125
14	56	0.2	53-67	52-69	51	120	91	111-125	110-126
15	57	0.2	54-68	53-70	52	121	92	112-126	111-127
16	59	0.3	56-70	55-71	53	122	93	113-127	111-128
17	61	0.5	58-72	57-73	54	124	95	115-129	113-130
18	63	1	60-74	58-75	55	125	95	115-130	114-131
19	65	1	61-76	60-77	56	127	96	117-131	116-133
20	66	1	62-76	61-78	57	128	97	118-132	117-134
21	68	2	64-78	63-80	58	130	98	120-134	119-135
22	70	2	66-80	65-81	59	131	98	121-135	120-136
23	71	3	67-81	66-82	60	132	98	122-136	120-137
24	73	4	69-83	67-84	61	134	99	124-138	122-139
25	74	4	70-84	68-85	62	135	99	124-139	123-140
26	75	5	70-85	69-86	63	137	99	126-140	125-142
27	76	5	71-85	70-87	64	139	99.5	128-142	127-143
28	78	7	73-87	72-89	65	140	99.6	129-143	128-144
29	79	8	74-88	73-89	66	142	99.7	131-145	129-146
30	81	10	76-90	75-91	67	143	99.8	132-146	130-147
31	83	13	78-92	76-93	68	145	99.9	133-148	132-149
32	85	16	79-94	78-95	69	146	99.9	134-148	133-150
33	87	19	81-95	80-97	70	148	99.9	136-150	135-152
34	89	23	83-97	82-98	71	150	>99.9	138-152	137-153
35	91	27	85-99	84-100	72	152	>99.9	140-154	138-155
36	93	32	87-101	85-102	73	154	>99.9	142-156	140-157
37	95	37	88-103	87-104	74	156	>99.9	143-157	142-159
38	97	42	90-104	89-106	75	158	>99.9	145-159	144-161
39	98	45	91-105	90-107	76	160	>99.9	147-161	146-162
40	100	50	93-107	92-108					

Interpretación básica del ICG y del ICC

Los pasos que figuran a continuación son una guía básica para comparar el ICG con el CI total y con el ICC a fin de obtener más información sobre el rendimiento de un niño en la WPPSI-IV. La interpretación del ICG y del ICC es opcional y se creó para complementar el análisis primario del perfil descrito en el capítulo 7 de este manual.

Evaluación del ICG y del ICC

El ICG y el ICC *siempre deberían analizarse junto* con el resto de las puntuaciones compuestas y de las puntuaciones de las pruebas. Las diferencias significativas e infrecuentes entre las puntuaciones de los índices primarios, así como la variabilidad excesiva entre las puntuaciones de las pruebas que contribuyen a dichos índices, pueden sugerir que el ICG o el ICC son una síntesis de diversas aptitudes.

Comparación y evaluación de las diferencias entre el ICG y el CI total y entre el ICG y el ICC

El procedimiento para comparar las diferencias entre las puntuaciones compuestas se describe detalladamente en el capítulo 2 del *Manual de aplicación y corrección*. Para comparar las diferencias entre el ICG y el CI total y entre el ICG y el ICC, debe seguirse el procedimiento descrito y utilizar las tablas C.4 y C.5 de este anexo, que permiten cumplimentar la tabla «Comparación entre índices/pruebas/puntuaciones de procesamiento» de la página de análisis secundario del cuadernillo de anotación. La tabla C.4 presenta los valores críticos de las diferencias estadísticamente significativas entre el ICG y el CI total y entre el ICG y el ICC, por grupo y franja de edad. La tabla C.5 presenta los porcentajes acumulados (tasas base) correspondientes a las diferencias entre el ICG y el CI total y entre el ICG y el ICC obtenidas por la muestra total y según el nivel de aptitud del ICG.

El primer paso de estas comparaciones es determinar si el valor absoluto de la diferencia es estadísticamente significativo. Cuando el valor absoluto de la diferencia es igual o superior al valor crítico, se considera una diferencia real y no debida a errores de medida o a efectos aleatorios. Si el ICG y el CI total no son significativamente diferentes, implica que reducir la influencia de la memoria de trabajo y de la velocidad de procesamiento en la estimación de la aptitud general tiene poco o ningún impacto sobre el rendimiento general del niño. Si el ICG y el ICC no son significativamente diferentes, implica que la competencia cognitiva es acorde con la capacidad general.

Si las comparaciones del ICG con el CI total y del ICG con el ICC indican una diferencia significativa, el profesional debe juzgar la frecuencia de esta diferencia respecto a la muestra de tipificación. Esta tasa base ofrece una estimación de la frecuencia de la diferencia entre las puntuaciones obtenidas por un niño en comparación con la población general (muestra total) o con niños con un nivel de aptitud similar (ICG).

Tabla C.4. Valores críticos para las diferencias estadísticamente significativas entre el ICG y determinadas puntuaciones compuestas, por grupo y franja de edad

Grupo/franja de edad	Nivel de significación	ICG-CIT	ICG-ICC
2:6–2:11	.01	4.88	
	.05	3.71	
	.10	3.10	
	.15	2.73	
3:0–3:5	.01	5.32	
	.05	4.04	
	.10	3.38	
	.15	2.97	
3:6–3:11	.01	4.61	
	.05	3.50	
	.10	2.93	
	.15	2.57	
4:0–4:5	.01	5.59	15.47
	.05	4.25	11.75
	.10	3.55	9.83
	.15	3.12	8.63
4:6–4:11	.01	5.90	18.16
	.05	4.48	13.79
	.10	3.75	11.54
	.15	3.29	10.13
5:0–5:5	.01	5.40	16.42
	.05	4.10	12.47
	.10	3.43	10.44
	.15	3.01	9.16
5:6–5:11	.01	6.22	16.86
	.05	4.73	12.81
	.10	3.96	10.72
	.15	3.47	9.41
6:0–6:11	.01	7.80	20.11
	.05	5.93	15.28
	.10	4.96	12.78
	.15	4.36	11.22
7:0–7:7	.01	6.39	16.41
	.05	4.85	12.46
	.10	4.06	10.43
	.15	3.56	9.18
2:6–3:11	.01	4.91	
	.05	3.73	
	.10	3.12	
	.15	2.74	
4:0–7:7	.01	6.21	17.30
	.05	4.72	13.15
	.10	3.95	11.00
	.15	3.47	9.66

Tabla C.5. Porcentajes acumulados (tasas base) de las diferencias entre el ICG y determinadas puntuaciones compuestas, a partir de toda la muestra de tipificación y según nivel de aptitud en el ICG

	Muestra total				ICG ≤ 79				80 ≤ ICG ≤ 89				
Diferencia	ICG < CIT (-)	ICG > CIT (+)	ICG < ICC (-)	ICG > ICC (+)	ICG < CIT (-)	ICG > CIT (+)	ICG < ICC (-)	ICG > ICC (+)	ICG < CIT (-)	ICG > CIT (+)	ICG < ICC (-)	ICG > ICC (+)	Diferencia
≥40	0.0	0.0	0.3	0.2	0.0	0.0	4.9	0.0	0.0	0.0	0.0	0.0	≥40
39	0.0	0.0	0.3	0.3	0.0	0.0	4.9	0.0	0.0	0.0	0.0	0.0	39
38	0.0	0.0	0.3	0.3	0.0	0.0	4.9	0.0	0.0	0.0	0.0	0.0	38
37	0.0	0.0	0.3	0.3	0.0	0.0	4.9	0.0	0.0	0.0	0.0	0.0	37
36	0.0	0.0	0.3	0.5	0.0	0.0	4.9	0.0	0.0	0.0	0.0	0.0	36
35	0.0	0.0	0.3	0.9	0.0	0.0	4.9	0.0	0.0	0.0	0.0	0.0	35
34	0.0	0.0	0.7	0.9	0.0	0.0	4.9	0.0	0.0	0.0	2.2	0.0	34
33	0.0	0.0	0.9	1.2	0.0	0.0	7.3	0.0	0.0	0.0	2.2	0.0	33
32	0.0	0.0	1.0	1.2	0.0	0.0	7.3	0.0	0.0	0.0	2.2	0.0	32
31	0.0	0.0	1.2	1.5	0.0	0.0	7.3	0.0	0.0	0.0	2.2	0.0	31
30	0.0	0.0	1.5	1.7	0.0	0.0	7.3	0.0	0.0	0.0	3.3	0.0	30
29	0.0	0.0	1.9	2.2	0.0	0.0	9.8	0.0	0.0	0.0	4.4	0.0	29
28	0.0	0.0	2.2	2.4	0.0	0.0	9.8	0.0	0.0	0.0	5.6	0.0	28
27	0.0	0.0	2.7	2.7	0.0	0.0	12.2	0.0	0.0	0.0	6.7	0.0	27
26	0.0	0.0	3.1	2.9	0.0	0.0	14.6	0.0	0.0	0.0	7.8	0.0	26
25	0.0	0.0	3.4	3.6	0.0	0.0	14.6	0.0	0.0	0.0	8.9	0.0	25
24	0.0	0.0	4.3	3.9	0.0	0.0	19.5	0.0	0.0	0.0	10.0	0.0	24
23	0.0	0.0	5.3	4.9	0.0	0.0	19.5	0.0	0.0	0.0	13.3	0.0	23
22	0.0	0.0	6.0	5.8	0.0	0.0	22.0	0.0	0.0	0.0	13.3	0.0	22
21	0.0	0.0	7.2	7.0	0.0	0.0	22.0	0.0	0.0	0.0	18.9	0.0	21
20	0.0	0.0	8.0	8.0	0.0	0.0	22.0	0.0	0.0	0.0	18.9	0.0	20
19	0.0	0.0	9.4	8.9	0.0	0.0	22.0	0.0	0.0	0.0	20.0	0.0	19
18	0.0	0.0	11.4	10.2	0.0	0.0	36.6	0.0	0.0	0.0	20.0	0.0	18
17	0.0	0.0	12.8	12.3	0.0	0.0	36.6	0.0	0.0	0.0	25.6	0.0	17
16	0.2	0.1	14.5	13.0	1.6	0.0	46.3	0.0	0.0	0.0	26.7	0.0	16
15	0.3	0.1	16.4	15.7	1.6	0.0	46.3	2.4	0.0	0.0	32.2	0.0	15
14	0.3	0.2	18.1	17.2	1.6	0.0	51.2	2.4	0.0	0.0	33.3	1.1	14
13	0.6	0.2	19.1	18.6	1.6	0.0	51.2	4.9	0.0	0.0	34.4	3.3	13
12	0.9	0.5	21.5	21.5	3.2	0.0	51.2	4.9	0.0	0.0	40.0	4.4	12
11	1.8	1.0	24.4	22.2	3.2	0.0	51.2	4.9	2.7	0.0	45.6	4.4	11
10	2.8	1.9	26.1	24.7	4.8	0.0	53.7	4.9	2.7	0.0	50.0	6.7	10
9	3.6	3.0	27.8	26.5	4.8	0.0	56.1	4.9	3.4	1.4	51.1	6.7	9
8	5.5	5.0	30.5	29.9	4.8	1.6	63.4	7.3	7.5	3.4	58.9	7.8	8
7	8.4	7.8	33.4	31.1	6.5	1.6	63.4	7.3	8.9	4.1	58.9	7.8	7
6	12.5	11.8	35.3	33.8	14.5	4.8	63.4	9.8	12.3	8.9	63.3	7.8	6
5	17.7	16.3	36.9	36.2	22.6	4.8	63.4	12.2	20.5	12.3	63.3	11.1	5
4	23.9	22.0	39.6	40.6	33.9	12.9	63.4	12.2	24.7	15.1	67.8	16.7	4
3	29.7	28.1	42.0	43.3	40.3	14.5	63.4	19.5	33.6	19.2	67.8	16.7	3
2	35.9	38.0	45.7	47.6	45.2	21.0	65.9	22.0	37.7	30.1	74.4	21.1	2
1	43.6	46.8	48.0	48.8	54.8	25.8	68.3	24.4	49.3	40.4	74.4	21.1	1
Media	*4.3*	*3.9*	*11.8*	*11.4*	*4.5*	*3.4*	*18.3*	*5.9*	*4.1*	*3.3*	*13.8*	*6.5*	*Media*
Dt	2.9	2.7	8.1	8.2	3.2	2.0	9.8	4.7	2.8	2.3	8.3	4.2	*Dt*
Mediana	4.0	3.0	11.0	10.0	4.0	3.5	18.0	4.0	3.5	2.0	12.0	5.0	Mediana

Tabla C.5. Porcentajes acumulados (tasas base) de las diferencias entre el ICG y determinadas puntuaciones compuestas, a partir de toda la muestra de tipificación y según nivel de aptitud en el ICG *(continuación)*

	90 ≤ ICG ≤ 109				110 ≤ ICG ≤ 119				ICG ≤ 120				
Diferencia	ICG < CIT (-)	ICG > CIT (+)	ICG < ICC (-)	ICG > ICC (+)	ICG < CIT (-)	ICG > CIT (+)	ICG < ICC (-)	ICG > ICC (+)	ICG < CIT (-)	ICG > CIT (+)	ICG < ICC (-)	ICG > ICC (+)	Diferencia
≥40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9	≥40
39	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.7	39
38	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.7	38
37	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.7	37
36	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6	36
35	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	7.4	35
34	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	7.4	34
33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	11.1	33
32	0.0	0.0	0.3	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	11.1	32
31	0.0	0.0	0.7	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	13.0	31
30	0.0	0.0	1.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	14.8	30
29	0.0	0.0	1.0	0.0	0.0	0.0	0.0	3.9	0.0	0.0	0.0	16.7	29
28	0.0	0.0	1.3	0.3	0.0	0.0	0.0	3.9	0.0	0.0	0.0	16.7	28
27	0.0	0.0	1.7	0.3	0.0	0.0	0.0	3.9	0.0	0.0	0.0	20.4	27
26	0.0	0.0	1.7	0.3	0.0	0.0	0.0	3.9	0.0	0.0	0.0	22.2	26
25	0.0	0.0	1.7	1.0	0.0	0.0	1.0	5.9	0.0	0.0	0.0	22.2	25
24	0.0	0.0	2.3	1.3	0.0	0.0	1.0	5.9	0.0	0.0	0.0	24.1	24
23	0.0	0.0	3.3	1.7	0.0	0.0	1.0	6.9	0.0	0.0	0.0	31.5	23
22	0.0	0.0	4.3	2.7	0.0	0.0	1.0	7.8	0.0	0.0	0.0	33.3	22
21	0.0	0.0	5.0	3.7	0.0	0.0	1.0	11.8	0.0	0.0	0.0	33.3	21
20	0.0	0.0	6.4	4.0	0.0	0.0	1.0	16.7	0.0	0.0	1.9	33.3	20
19	0.0	0.0	8.4	5.7	0.0	0.0	2.0	16.7	0.0	0.0	1.9	33.3	19
18	0.0	0.0	10.0	7.0	0.0	0.0	2.0	19.6	0.0	0.0	3.7	35.2	18
17	0.0	0.0	10.7	8.7	0.0	0.0	2.9	24.5	0.0	0.0	3.7	38.9	17
16	0.2	0.0	12.4	8.7	0.0	0.0	2.9	27.5	0.0	1.2	3.7	40.7	16
15	0.5	0.0	14.0	11.4	0.0	0.0	3.9	30.4	0.0	1.2	3.7	48.1	15
14	0.5	0.0	15.4	12.7	0.0	0.0	4.9	32.4	0.0	2.4	7.4	51.9	14
13	0.9	0.0	17.1	14.0	0.0	0.0	4.9	32.4	0.0	2.4	7.4	53.7	13
12	1.1	0.0	19.1	16.1	0.0	0.7	6.9	39.2	1.2	3.6	9.3	59.3	12
11	1.8	0.2	22.1	17.4	0.7	1.4	8.8	39.2	1.2	7.2	11.1	59.3	11
10	2.5	0.9	23.7	20.4	3.4	2.7	8.8	43.1	2.4	10.8	11.1	59.3	10
9	3.4	2.3	26.4	22.7	4.8	2.7	8.8	43.1	2.4	12.0	11.1	64.8	9
8	4.8	3.2	27.8	26.1	6.8	6.8	10.8	50.0	3.6	16.9	11.1	66.7	8
7	8.6	6.1	32.8	27.4	10.3	13.7	11.8	52.0	4.8	18.1	13.0	68.5	7
6	12.0	9.0	35.1	30.8	16.4	18.5	11.8	54.9	7.2	25.3	13.0	72.2	6
5	17.0	14.7	37.8	32.1	19.2	20.5	12.7	58.8	10.8	32.5	13.0	75.9	5
4	23.8	20.8	41.8	37.5	25.3	27.4	12.7	61.8	13.3	37.3	13.0	79.6	4
3	29.9	27.4	45.2	39.1	28.1	34.2	15.7	69.6	16.9	47.0	14.8	79.6	3
2	35.7	37.8	48.8	44.5	37.7	43.8	19.6	72.5	24.1	55.4	14.8	81.5	2
1	43.0	47.5	51.5	46.2	41.1	51.4	23.5	73.5	32.5	61.4	14.8	81.5	1
Media	4.3	3.6	10.3	9.6	4.7	4.4	7.7	12.5	3.7	5.5	12.4	18.3	Media
Dt	2.9	2.3	7.2	6.5	2.8	2.7	6.8	8.0	2.9	3.6	5.5	10.5	*Dt*
Mediana	4.0	3.0	9.0	8.0	4.0	4.0	6.0	12.0	3.0	5.0	13.0	15.5	*Mediana*

Criterios de inclusión para los grupos especiales

Criterios de inclusión generales para los grupos especiales (muestra estadounidense)

Los niños fueron seleccionados para formar parte de la muestra de los grupos especiales si cumplían los siguientes criterios:

- inglés como lengua materna;
- capacidad para comunicar a un nivel acorde con la edad y el diagnóstico, y cierta capacidad para comunicar oralmente;
- audición y visión normales (corregidas);
- motricidad gruesa y fina normal (a excepción del déficit motriz leve que se da en los grupos de niños con discapacidad intelectual y con retraso del desarrollo cognitivo);
- ausencia de enfermedad o condición física que pudiera afectar al funcionamiento cognitivo o al rendimiento en las pruebas (a excepción de los trastornos asociados al grupo especial al que pertenece el niño);
- no haber sido diagnosticado con una afección neurológica (p. ej., trastorno convulsivo, epilepsia, encefalitis, cirugía cerebral, tumor cerebral) distinta a la pertinente al grupo especial correspondiente;
- no haber presentado un periodo de inconsciencia de 20 minutos o más (a excepción del grupo de sujetos con factores de riesgo para el desarrollo);
- no haber sido diagnosticado con discapacidad intelectual (a excepción del grupo con discapacidad intelectual) ni con trastorno generalizado del desarrollo (a excepción de los grupos de niños con trastorno autista y con trastorno de Asperger);
- no haber sido diagnosticado con un trastorno psiquiátrico (p. ej., trastornos del lenguaje, trastornos psicóticos, trastornos del estado de ánimo) distinto a los establecidos como criterios de los grupos especiales;
- no estar internado en un hospital o en una institución mental o psiquiátrica, ni seguir ningún tratamiento (a excepción de los ámbitos en que se siguen tratamientos específicos para una afección determinada [p. ej., centros para los sujetos con discapacidad intelectual o trastorno autista];
- no estar tomando en ese momento ninguna medicación que pudiera afectar al rendimiento del niño en la escala, a excepción de la adecuada a la afección que define al grupo especial, o a otras afecciones asociadas;
- cumplir los criterios correspondientes exclusivamente a uno de los grupos clínicos (a excepción de los diagnósticos duales, como el trastorno por déficit de atención con hiperactividad o el comportamiento perturbador secundarios a la discapacidad intelectual); y
- no haber sido evaluado con la WPPSI-III u otra medida de la aptitud cognitiva durante los últimos 5 meses antes de la fecha de aplicación de la escala.

Criterios de inclusión específicos para los grupos especiales (muestra estadounidense)

Alta capacidad intelectual

Criterios:

- tener entre 4:0 y 7:7 años

Y

- puntuar ≥ 2 desviaciones típicas por encima de la media en pruebas estandarizadas de aptitud intelectual aplicadas individualmente (p. ej., CI ≥ 130).

O

- *recibir atención especial para la alta capacidad intelectual, si no se dispone de puntuaciones.*

Discapacidad intelectual

Criterios:

- tener entre 2:6 y 7:7 años

Y

- cumplir los criterios del DSM-IV-TR para el diagnóstico de discapacidad intelectual leve o moderada;

O

- puntuar entre 2 y 4 desviaciones típicas por debajo de la media en pruebas estandarizadas de aptitud intelectual aplicadas individualmente (p. ej., CI = 40-70).

Retraso del desarrollo cognitivo

Criterios:

- tener entre 2:6 y 7:7 años

Y

- haber sido identificado con un trastorno del desarrollo en el área cognitiva.

O

- puntuar entre 1 y 2 desviaciones típicas por debajo de la media en pruebas estandarizadas de aptitud intelectual aplicadas individualmente (p. ej., CI = 71-85).

Factores de riesgo para el desarrollo

Criterios:

- tener entre 2:6 y 7:7 años;
- *haber sido identificado con factores de riesgos biológicos o ambientales asociados a una mayor probabilidad de presentar un retraso del desarrollo o un retraso cognitivo (p. ej., anomalías cromosómicas, desnutrición, síndrome alcohólico fetal, dificultad respiratoria perinatal, hemorragia cerebral, bajo peso al nacer); y*
- no estar basados los factores de riesgo en factores sociales o nivel socioeconómico.

Problemas previos a la lectoescritura

Criterios:

- tener entre 4:6 y 7:7 años;
- puntuar ≥ 85 en la *Wechsler Individual Achievement Test (3.ª edición)*.

Trastorno por déficit de atención con hiperactividad

Criterios:

- tener entre 3:0 y 7:7 años;
- puntuar ≥ 80 en pruebas estandarizadas de aptitud intelectual aplicadas individualmente, o disponer de un informe psicológico en el que se estima un CI ≥ 80;
- cumplir los criterios del DSM-IV-TR para el diagnóstico del trastorno por déficit de atención con hiperactividad;
- presentar resultados clínicamente significativos de las *Brown Attention-Deficit Disorder Scales for Children and Adolescents* (Brown, 2001) en su forma para los padres; y
- no haber tomado medicación psicoestimulante durante, como mínimo, 24 horas antes de la aplicación de la escala.

Comportamiento perturbador

Criterios:

- tener entre 4:0 y 7:7 años;
- puntuar ≥ 80 en pruebas estandarizadas de aptitud intelectual aplicadas individualmente, o disponer de un informe psicológico en el que se estima un CI ≥ 80;
- puntuar ≥ 65 en la escala Autocontrol emocional, en su forma para los padres, de la *Behavior Assessment System for Children-Second Edition* (Reynolds y Kamphaus, 2004);

Y

- cumplir los criterios del DSM-IV-TR para el diagnóstico del trastorno negativista desafiante o del trastorno de la conducta.

O

- haber sido clasificado como que presentaba un trastorno emocional, en función de los criterios del centro escolar y ante su comportamiento perturbador.

Trastorno del lenguaje expresivo

Criterios:

- tener entre 4:0 y 7:7 años;
- puntuar ≥ 80 en pruebas estandarizadas de aptitud intelectual aplicadas individualmente, o disponer de un informe psicológico en el que se estima un CI ≥ 80;

Y

- cumplir los criterios del DSM-IV-TR para el diagnóstico del trastorno del lenguaje expresivo;

O

- presentar una puntuación del lenguaje expresivo dentro del rango clínicamente significativo en una medida global del lenguaje, tal y como se indica en el manual correspondiente.

Trastorno mixto del lenguaje receptivo-expresivo

Criterios:

- tener entre 4:0 y 7:7 años;
- puntuar ≥ 80 en pruebas estandarizadas de aptitud intelectual aplicadas individualmente, o disponer de un informe psicológico en el que se estima un CI ≥ 80;

Y

- cumplir los criterios del DSM-IV-TR para el diagnóstico del trastorno mixto del lenguaje receptivo-expresivo;

O

- presentar una puntuación tanto del lenguaje expresivo como del receptivo dentro del rango clínicamente significativo en una medida global del lenguaje, tal y como se indica en el manual correspondiente.

Trastorno autista

Criterios:

- tener entre 2:6 y 7:7 años;
- cumplir los criterios del DSM-IV-TR para el diagnóstico del trastorno autista;
- contar con las destrezas lingüísticas para realizar el test; y
- puntuar ≥ 60 en pruebas estandarizadas de aptitud intelectual aplicadas individualmente, o disponer de un informe psicológico en el que se estima un CI ≥ 60.

Trastorno de Asperger

Criterios:

- tener entre 3:6 y 7:7 años;
- cumplir los criterios del DSM-IV-TR para el diagnóstico del trastorno de Asperger;
- no presentar retrasos del lenguaje clínicamente significativos; y
- puntuar ≥ 70 en pruebas estandarizadas de aptitud intelectual aplicadas individualmente, o disponer de un informe psicológico en el que se estima un CI ≥ 70.

WPPSI-IV

ESCALA DE INTELIGENCIA DE WECHSLER
PARA PREESCOLAR Y PRIMARIA-IV

ISBN 978-84-1555-233-8
9 788415 552338

PEARSON PsychCorp Pearson Clinical & Talent Assessment www.pearsonpsychcorp.es