

Estandarización Chilena de la Escala Wechsler de Inteligencia para Niños—Quinta Edición

Chilean Standardization of the Wechsler Intelligence Scale for Children—Fifth Edition

Ricardo Rosas¹, Marcelo Pizarro², Olivia Grez², Valentina Navarro², Dolly Tapia³,
Susana Arancibia⁴, María Teresa Muñoz-Quezada⁵, Boris Lucero⁵,
Claudia P. Pérez-Salas⁶, Karen Oliva⁶, Beatriz Vizcarra⁷,
Marcela Rodríguez-Cancino⁷ y Patricia von Freeden⁸

¹ Escuela de Psicología y Centro de Desarrollo de Tecnologías de Inclusión,
Pontificia Universidad Católica de Chile

² Centro de Desarrollo de Tecnologías de Inclusión, Pontificia Universidad Católica de Chile

³ Departamento de Administración de Educación Municipal, Arica, Chile

⁴ Centro de Innovación Metodológica y Tecnológica, Universidad Católica del Norte

⁵ Centro de Investigación en Neuropsicología y Neurociencias Cognitivas,
Universidad Católica del Maule

⁶ Departamento de Psicología, Universidad de Concepción

⁷ Departamento de Psicología, Universidad de La Frontera

⁸ Facultad de Psicología, Universidad San Sebastián

La Escala Wechsler de Inteligencia para Niños — Quinta Edición es la más reciente actualización de estas escalas. Esta nueva edición fue lanzada el año 2014 en Estados Unidos y presenta mejoras relevantes respecto de versiones anteriores, como una estructura factorial actualizada que incluye un mayor número de indicadores, nuevos puntajes de proceso y nuevos procedimientos para la construcción de las normas. Estas mejoras son relevantes en el contexto nacional, puesto que la versión que estaba previamente disponible en Chile era WISC-IIIv.ch, cuya estructura factorial data del año 1991, aunque fue estandarizada el año 2007 para la población chilena. En este artículo se presenta el procedimiento de estandarización de WISC-V para la población chilena. Este procedimiento incluye la traducción, adaptación y construcción de normas. Se presenta evidencia de validez y confiabilidad, así como también los resultados de la aplicación del instrumento a muestras de grupos especiales, como discapacidad intelectual (24), trastorno de déficit atencional con hiperactividad (27), trastorno del espectro autista (26), trastorno específico del lenguaje (56) y población rural (47). Los resultados dan cuenta de excelentes resultados de confiabilidad, así como de un excelente ajuste al modelo factorial utilizado para la versión norteamericana. Por último, se presenta evidencia de validez convergente entre WISC-V y otras versiones de estas escalas de inteligencia, como WAIS-IV y WISC-IIIv.ch. También se presentan comparaciones de rendimiento entre los tres tipos de dependencia educativa existentes en Chile, así como también de comparaciones entre género. Los resultados de esta investigación dan cuenta que la

Ricardo Rosas  <https://orcid.org/0000-0002-3091-4044>

Marcelo Pizarro  <https://orcid.org/0000-0003-2333-804X>

Olivia Grez  <https://orcid.org/0000-0003-2027-3439>

Valentina Navarro  <https://orcid.org/0000-0002-2382-2731>

Susana Arancibia  <https://orcid.org/0000-0002-2382-2731>

María Teresa Muñoz-Quezada  <https://orcid.org/0000-0002-8008-8625>

Boris Lucero  <https://orcid.org/0000-0002-7081-3344>

Claudia P. Pérez-Salas  <https://orcid.org/0000-0002-6940-6514>

Karen Oliva  <https://orcid.org/0000-0003-1220-7570>

Beatriz Vizcarra  <https://orcid.org/0000-0002-7052-7882>

Marcela Rodríguez-Cancino  <https://orcid.org/0000-0003-2012-9701>

La correspondencia relativa a este artículo debe ser dirigida a Ricardo Rosas, Escuela de Psicología, Pontificia Universidad Católica de Chile, Avda. Vicuña Mackenna 4860, Macul, Santiago, Región Metropolitana, Chile. Email: rrosas@uc.cl

estandarización chilena de WISC-V cuenta con excelentes características psicométricas que hacen recomendable su uso en el país.

Palabras clave: evaluación, inteligencia, estandarización, WISC-V, Chile

The Wechsler Intelligence Scale for Children — Fifth Edition is the most recent update to this scale, launched originally in 2014 in the United States. WISC-V includes several improvements from previous versions, such as an updated factorial structure that includes more indexes, new process scores and new procedures for norming. These improvements are particularly relevant in the Chilean context, since the previous version available in the country was the WISC-IIIv.ch, whose factorial structure dates from 1991, although it was standardized for Chilean population in 2007. This article presents the standardization process for the WISC-V, which includes the translation, adaptation and norming procedures for the Chilean population. It includes evidence for reliability and validity, and also results for various special groups samples, such as intellectual disability (24), attentional deficit disorder with hyperactivity (27), autism spectrum disorder (26), specific language disorder (56) and rural population (47). The Chilean data shows excellent reliability results and an excellent adjustment to the factorial model used in the North American version. This research provides evidence for convergent validity between WISC-V and other versions of these scales, such as the WAIS-IV and the WISC-IIIv.ch. Finally, we perform a comparison between private, voucher and public schools using this scale and a comparison between male and female performance in this scale. The main finding of this research is that the Chilean standardization of the WISC-V has excellent psychometric properties that allow recommending the use of this instrument in the country.

Keywords: assessment, intelligence, standardization, WISC-V, Chile

Sobre la Necesidad de una Actualización de la Escala Wechsler de Inteligencia para Niños

Las Escalas Wechsler de Inteligencia son una de las baterías que se aplican con más frecuencia en el ámbito de la evaluación cognitiva. Estas escalas son consideradas el estándar mundial, dada su historia y utilidad clínica. En Chile, el Ministerio de Educación, a través del decreto 170 (Fija normas para determinar los alumnos, 2009), recomienda el uso de estas escalas en su versión más actualizada como uno de los insumos necesarios para tomar la decisión de ingreso al Programa de Integración Escolar (PIE) de cada colegio. En estos programas, los estudiantes que presentan Necesidades Educativas Especiales (NEE), ya sea de tipo permanente o transitorio, reciben apoyos por parte de varios especialistas, como psicólogos, fonoaudiólogos o psicopedagogos, entre otros.

En este artículo se describe el proceso de adaptación y estandarización de WISC-V para la población chilena. Esta actualización resulta necesaria para la evaluación cognitiva en Chile, dado que la versión anterior de la batería disponible, WISC-IIIv.ch, no considera los avances realizados durante los últimos 20 años en teorías de evaluación, cognición, aprendizaje y desarrollo. A continuación, se presenta una breve historia de las escalas Wechsler disponibles en Chile.

Evolución de las Escalas Wechsler y su Disponibilidad en Chile

La primera escala Wechsler infantil disponible en Chile fue una adaptación de WISC original, realizada por Edda Campazzo (1962) de la Pontificia Universidad Católica de Chile en 1960, aunque sin datos de estandarización. Posteriormente, se adaptó y estandarizó WISC-R entre los años 1975 y 1980 por tres grupos de tesis de la misma universidad, que recogieron datos únicamente en Santiago (Adriasola et al., 1976; Calderón et al., 1980; Cañas et al., 1978). Estas investigaciones no se materializaron en un producto construido en forma estandarizada, por lo que solamente se vendían copias cuyas calidades de producción variaban. La estructura factorial de este instrumento tenía dos niveles de análisis: un primer nivel de Escala Total, que daba cuenta de la capacidad cognitiva del individuo y un segundo nivel de escalas en el que estaban las así llamadas escalas Verbal y Manual.

Posteriormente, entre los años 2002 y 2003, se recogieron datos normativos para el lanzamiento de WISC-IIIv.ch. Esta batería marcó un hito en la cultura de evaluación chilena, pues era la primera vez que uno de estos instrumentos se producía de manera legal, autorizada por el editor internacional de la batería y con procedimientos de producción estandarizados en el país. Esta herramienta fue adaptada y estandarizada en el marco de un proyecto del Fondo Nacional de Desarrollo Científico y Tecnológico, FONDECYT 1020951 (Ramírez & Rosas, 2007). WISC-IIIv.ch tenía una estructura de análisis de tres niveles: el primer nivel correspondiente al indicador general de la batería (CIT), el segundo nivel correspondiente a las Escalas Verbal y Ejecutiva y el tercer nivel que incluía cuatro nuevos subíndices, Comprensión Verbal (ICV),

Organización Perceptual (OP), Velocidad de Procesamiento (IVP) y Ausencia de Distractibilidad (AD), que permitían realizar evaluaciones con un mayor grado de discriminación entre los distintos constructos.

WISC-IV salió al mercado el año 2003 en Estados Unidos. Esta versión de la batería tuvo cambios relevantes en cuanto a la estructura y el proceso de desarrollo. Respecto de la estructura, el mayor cambio fue la eliminación del nivel de análisis entre las Escalas Verbal y Ejecutiva, puesto que resultaba demasiado amplio para ser útil clínicamente. En consecuencia, la estructura de WISC-IV tenía dos niveles: un primer nivel de Escala Total en el que se incluía el CIT y un segundo nivel en el que se incluían cuatro índices, Compresión Verbal (ICV), Razonamiento Perceptual (IRP), Velocidad de Procesamiento (IVP) y Memoria de Trabajo (IMT). La variación en los nombres de los índices respecto de WISC-III (e.g., desde AD a IMT) significó cambios en las subpruebas y composición de los índices para obtener mejores medidas de los constructos evaluados. En cuanto al proceso de desarrollo, la modificación más importante fue la incorporación de la teoría neuropsicológica, que se materializó en la inclusión de variados puntajes de proceso que permitieron abordar otros aspectos de la cognición. Además, esta fue una de las primeras escalas Wechsler que incluyó el procedimiento de normas inferenciales (Wechsler, 2003; Zhu & Chen, 2011), que permite obtener normas de mejor calidad con un menor número de casos; además, permite corregir ciertas anomalías en las normas por características atribuibles a la muestra, lo que mejora significativamente la calidad de los baremos y aumenta la utilidad clínica del instrumento. Es importante notar que el año 2003 - el mismo año en que estaban recogiendo datos para la versión chilena de WISC-III- salió al mercado en Estados Unidos WISC-IV. Es por esta razón que en Chile no se estandarizó WISC-IV y que se haya pasado directamente desde la tercera edición a la quinta.

El lanzamiento de WISC-V en Estados Unidos el año 2014 incluyó cambios importantes respecto de WISC-IV. La nueva versión de la batería se caracterizó por una estructura de cuatro niveles: Escala Total, Índices Principales, Índices Auxiliares e Índices Complementarios. Esta estructura permitió una mayor diferenciación del funcionamiento cognitivo que se materializó en los 14 índices que componen la versión norteamericana. En la misma línea, los análisis que se pueden realizar con esta batería son más sofisticados que los de sus versiones anteriores. Por ejemplo, una de las posibilidades de análisis es la comparación entre el IMT y el de Memoria de Trabajo Auditiva (IMTA), lo que permite evaluar si el tipo de estímulo -verbal o auditivo- influye en el rendimiento. Este instrumento, al igual que su versión anterior, también incluyó el procedimiento inferencial para la construcción de las normas.

Descripción del Instrumento

WISC-V es una prueba de administración individual que permite la evaluación de la inteligencia en niños y adolescentes desde los 6 años 0 meses hasta los 16 años 11 meses. La estructura de la prueba consta de tres niveles de análisis, entre los cuales están repartidos los 11 índices que componen la batería. En el nivel de escala total se incluye el indicador de Coeficiente Intelectual Total (CIT); en el nivel de índices principales se incluyen: ICV, Visoespacial (IVE), Razonamiento Fluido (IRF), IMT e IVP; en el nivel de índices secundarios se incluyen: IMTA, Razonamiento Cuantitativo (IRC), No Verbal (INV), Habilidad General (IHG) y Competencia Cognitiva (ICC). Entre estos 11 índices se reparten las 15 subpruebas incluidas en la batería. En la Figura 1 se resume la composición de los niveles de análisis y de los índices. Es importante notar que la versión estandarizada en Chile no incluye la subprueba Conceptos Visuales (*Picture Concept*) ni tampoco el nivel de análisis de los índices complementarios, en el que se incluyen tres índices adicionales. Estas subpruebas y niveles de análisis no fueron licenciados por Pearson para versiones fuera de Estados Unidos.

Método y Resultados

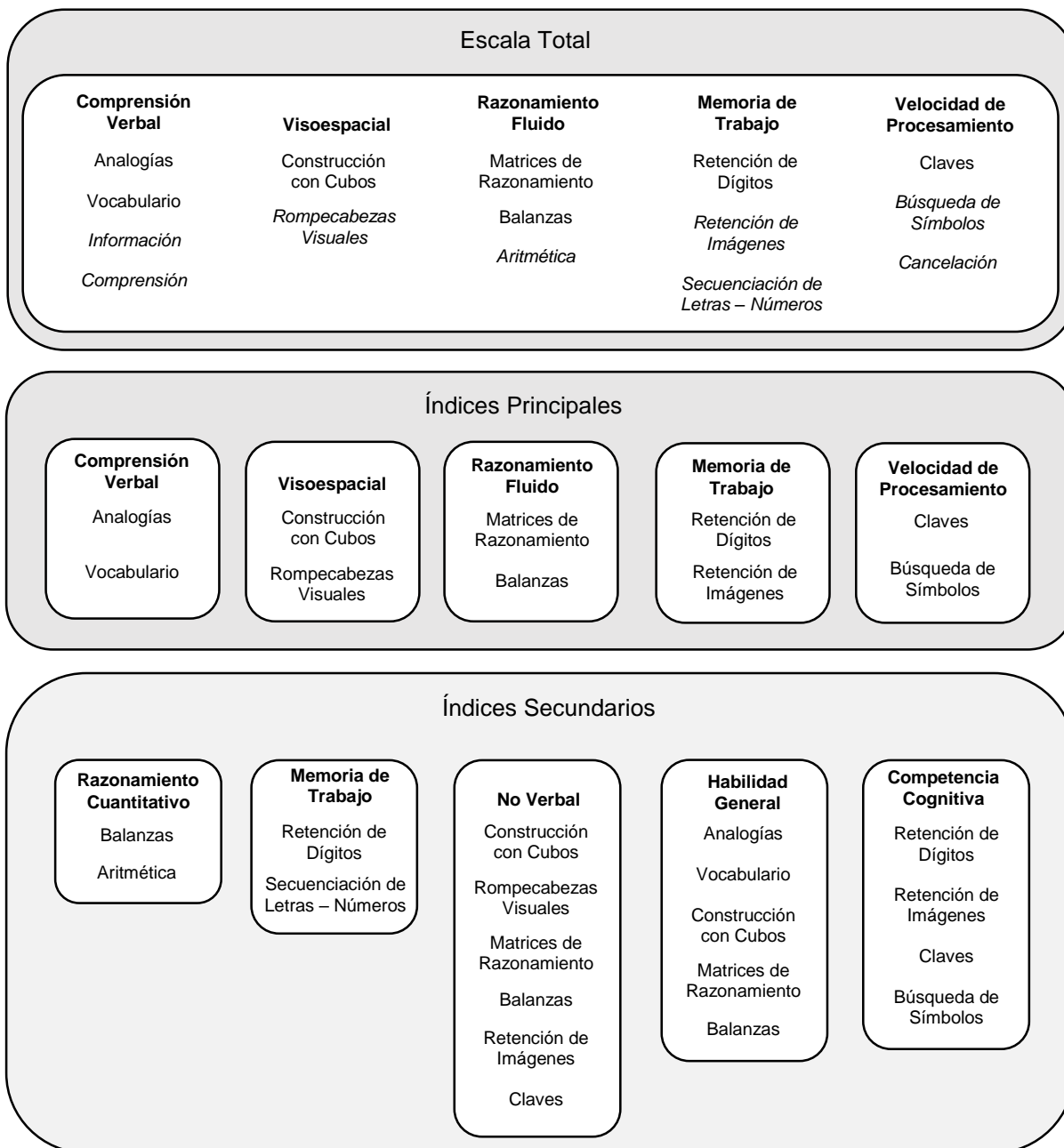
Adaptación y Estandarización

El proceso de adaptación y estandarización de WISC-V se realizó en cuatro etapas: (a) traducción y adaptación de materiales e ítems, (b) fase pre-piloto, (c) fase piloto y (d) fase de tipificación. A continuación se describen los objetivos, procedimientos y participantes (cuando corresponde) de cada una de estas etapas.

Es importante notar que el procedimiento completo de la estandarización del instrumento fue revisado y aprobado por el Comité Ético Científico en Ciencias Sociales, Artes y Humanidades de la Pontificia Universidad Católica de Chile. Previo a la aplicación del instrumento, se envió el consentimiento informado a los apoderados de cada uno de los niños, niñas y adolescentes que fueron invitados a participar en el estudio.

Solo fueron evaluados aquellos sujetos cuyos padres autorizaron la evaluación mediante la firma del consentimiento informado.

Figura 1
Estructura de WISC V



Nota. Las cursivas indican las subpruebas que pueden ser utilizadas como reemplazo para cada uno de los dominios incluidos para obtener el indicador de la Escala Total.

En caso de que la toma de datos se realizara en horario de clases, se coordinó con el establecimiento educacional un horario que no afectara los aprendizajes de los estudiantes. Llegado el momento de la aplicación, se leyó a cada participante un asentimiento informado en el que se les comunicaba sobre la investigación. Finalizado, se entregó a cada niña, niño y adolescente una pequeña retribución (una pegatina o un lápiz) como una forma de retribución por su tiempo dedicado a contestar el instrumento.

No se entregaron resultados individualizados ni grupales a ningún participante de la investigación. Los datos fueron almacenados en una base de datos única, en la que fueron anonimizados los datos de identificación de los participantes. Solo se conservaron los datos relevantes para el muestreo, que fueron la edad, sexo, tipo de establecimiento educacional y región del país. La base de datos solo fue vista por personas que formaban parte del equipo líder del proyecto.

Es preciso mencionar que en este artículo no se hará referencia a ningún ítem de la batería, puesto que es necesario resguardarlos para impedir su divulgación y, de esta manera, mantener la validez del instrumento.

Traducción y Adaptación de Materiales e Ítems

La primera etapa de desarrollo consistió en una traducción inicial de todos los materiales (Manual de Administración y Corrección, Protocolos de Registros para la Aplicación y Resultados y Cuadernillos de Respuesta). Posteriormente, se realizó una revisión de los ítems del ICV, que incluyó las subpruebas de Analogías (AN), Vocabulario (VOC), Información (INF) y Comprensión (COM). Se conservaron todos los ítems que eran pertinentes culturalmente para usarlos en el eje pacífico de América Latina. Los ítems que hacían alusión a elementos específicos de la cultura estadounidense fueron removidos, como aquellos referidos a la carrera espacial u otros relacionados al contexto educacional norteamericano. Además de lo anterior, se privilegiaron los ítems que aparecían en la versión norteamericana de WISC-V y en WISC-IIIv.ch para mantener una continuidad en las versiones del país. Es importante notar que, si un ítem aparecía en ambas versiones, pero existían diferencias en el fraseo entre una y otra, se conservó el fraseo de la versión de WISC-V.

La adaptación que se hizo a la subprueba VOCo requiere de atención especial. Para su adaptación se analizó cada una de las palabras de la versión norteamericana de la batería, utilizando el Corpus of Contemporary American English (Davies, 2008) para identificar la frecuencia de aparición y registro de origen. Posteriormente, se utilizó el Lifcach (Sadowsky & Martínez-Gamboa, 2012) para buscar palabras en el español de Chile que tuvieran una frecuencia y registro de origen similar a la versión norteamericana. La selección final de palabras fue revisada por profesionales chilenos, ecuatorianos, venezolanos y colombianos para lograr un español tan neutro como fuera posible. Por último, se agregaron ítems adicionales para tener un repertorio suficiente en caso de que algunos de los seleccionados no cumplieran con los criterios psicométricos necesarios para ser incluidos en la versión de tipificación. La traducción de todos los ítems verbales al español fue posteriormente retraducida al inglés y enviada a Pearson para su análisis y aprobación.

No se realizaron adaptaciones en las subpruebas no verbales, puesto que los ítems incluidos tienen una menor sensibilidad cultural.

Fase Pre-Piloto

Esta fase tuvo dos objetivos: (a) probar las instrucciones e ítems de ejemplo de todas las subpruebas y (b) probar inicialmente el fraseo de los ítems.

Participantes. En esta etapa se aplicó una versión prepiloto del instrumento a una muestra intencionada de 26 sujetos entre los 5 a 16 años de dos establecimientos educacionales municipales de la Región Metropolitana durante los meses de septiembre a diciembre de 2015. En esta muestra se incluyó a un participante con el diagnóstico de discapacidad intelectual leve y a otro con el diagnóstico de trastorno específico del lenguaje (TEL); los participantes restantes tenían un desarrollo cognitivo típico. La distribución de la muestra se presenta en la Tabla 1.

Resultados. En general, todos los estudiantes pudieron realizar las tareas sin inconvenientes. En el único ítem en el que se observaron ciertas dificultades fue en el primer ítem de práctica de la subprueba AN; aunque estas fueron resueltas con algunas preguntas adicionales, se decidió agregar un segundo ítem de práctica en esta subprueba para mejorar la comprensión de las instrucciones.

Tabla 1
Composición de la Muestra del Pre-Piloto

Edad	Composición de la muestra			Total
	Desarrollo típico	Discapacidad intelectual	Trastorno específico del lenguaje	
5	1	-	-	1
6	2	-	1	3
7	1	-	-	1
8	2	-	-	2
9	2	-	-	2
10	2	-	-	2
11	2	1	-	3
12	3	-	-	3
13	2	-	-	2
14	3	-	-	3
15	2	-	-	2
16	2	-	-	2
Total	24	1	1	26

Fase Piloto

Esta fase tuvo dos objetivos: (a) obtener información psicométrica de las subpruebas verbales y (b) obtener datos de confiabilidad de las subpruebas Claves (CLA), Búsqueda de Símbolos (BS) y Cancelación (CAN), dado que en estas subpruebas se requiere del procedimiento test-retest para evaluar la confiabilidad.

Participantes. En esta fase se recogieron datos de una muestra intencionada de 125 niños y adolescentes (63 hombres y 62 mujeres) entre 6 y 16 años ($M = 10,4$, $DE = 3,06$) de la Región Metropolitana, estratificados por dependencia escolar como proxy del nivel socioeconómico (NSE): municipal = 39, particular subvencionado = 49, particular pagado = 37. Es importante notar que los colegios municipales son financiados con recursos públicos, los particulares subvencionados, con recursos públicos y privados, y los particulares pagados, con recursos privados únicamente. Todos los participantes en esta etapa tenían un desarrollo cognitivo típico.

Instrumento. En esta etapa se aplicaron las siguientes subpruebas: AN, VOC, INF, COM, CLA, BS y CAN.

Procedimiento. A los apoderados de cada participante se les envió, a través del establecimiento educacional, una invitación a participar en el proyecto y un consentimiento informado. Solo se evaluaron aquellos sujetos cuyos apoderados aprobaron la participación mediante la firma del consentimiento informado.

En esta fase, que se realizó desde septiembre a diciembre de 2016, se aplicaron las subpruebas verbales completas: AN, VOC, INF y COM. No se aplicó el criterio de suspensión con el fin de obtener información acerca del nivel de dificultad de todos los ítems. Entre una y otra subprueba se aplicaron las subpruebas CLA, BS y CAN. Estas subpruebas se aplicaron una segunda vez al término de la evaluación con el objetivo de estimar la confiabilidad por estabilidad temporal mediante el procedimiento de medidas repetidas.

Análisis de Datos. Para identificar el nivel de dificultad de los ítems de las subpruebas verbales se realizó un análisis de frecuencia de respuestas. Para obtener datos de consistencia interna en estas subpruebas se utilizó el coeficiente alfa de Cronbach; en las subpruebas de CLA, BS y CAN se utilizó el procedimiento de medidas repetidas para obtener información sobre confiabilidad. Se utilizó el software SPSS 25 para realizar el análisis de datos.

Resultados. Los resultados de estos análisis se encuentran en la Tabla 2.

Tabla 2
Consistencia Interna y Correlación de Medidas Repetidas del Piloto

Subprueba	Sigla	Alfa de Cronbach	Correlación de medidas repetidas
Analogías	AN	0,900	-
Vocabulario	VOC	0,868	-
Información	INF	0,910	-
Comprensión	COM	0,850	-
Claves	CLA	-	0,935
Búsqueda de Símbolos	BS	-	0,810
Cancelación	CAN	-	0,862

Tipificación

Esta etapa tuvo el objetivo de construir la norma nacional. Las aplicaciones en regiones fueron realizadas por entidades estatales y académicas: Departamento de Administración de Educación Municipal de Arica, Universidad Católica del Norte en Antofagasta, Universidad Católica del Maule en Talca, Universidad de Concepción, Universidad de La Frontera en Temuco y la Universidad San Sebastián en Puerto Montt. Se incluyeron 14 casos (1,8% de la muestra) de personas con rendimientos similares a discapacidad intelectual para poder tener evidencia de los niveles bajos de desempeño. Estas evaluaciones fueron realizadas por profesionales del Hospital San Borja Arriarán. Para la muestra de tipificación también se incluyeron 47 casos de muestra rural (6,2% de la muestra).

Participantes. Para la construcción de normas se aplicó la batería a una muestra intencionada de 754 sujetos (387 mujeres y 367 hombres) entre 6 a 16 años ($M = 10,88$, $DE = 3,017$) estratificados por NSE (municipal = 286, particular subvencionado = 245 y particular pagado = 223) y repartida en siete zonas del país: Arica y Parinacota, Antofagasta, Región Metropolitana, Maule, Biobío, Araucanía y Los Lagos. Para la elección de estas zonas se tomó en consideración el que hubieran centros urbanos y universidades que previamente participaron en procesos de estandarización de instrumentos similares a WISC-V. La selección de los establecimientos educacionales se realizó en función de los resultados del Sistema de Medición de la Calidad de los Aprendizajes (SIMCE) de los años 2011, 2012 y 2013. El SIMCE es un conjunto de pruebas que se utilizan para evaluar los resultados de los establecimientos de educación primaria y secundaria. Se escogieron establecimientos educacionales de cada región que tuvieran resultados SIMCE que fueran lo más parecidos posible al promedio regional por dependencia. De esta manera, se buscó que los establecimientos seleccionados fueran lo más representativos posible tanto de la región como del país. En la Tabla 3 se presenta la composición de la muestra normativa.

Se envió a cada establecimiento educacional una lista con variables de exclusión, en la que se incluían distintos tipos de restricciones para los participantes (e.g., que tuviera un diagnóstico clínico a la fecha, que fuera familiar de los aplicadores o que tuviera hermanos que hubiesen participado en este proyecto o que tuvieran evaluaciones cognitivas similares a la fecha). Posteriormente, los establecimientos enviaron una lista con los estudiantes que podrían ser evaluados. De esta lista, se seleccionaron aquellos estudiantes que a la fecha de evaluación tuvieran edades que fueran centrales para los grupos etarios (e.g., se evitó seleccionar a participantes que estuvieran a punto de cumplir año) para que fueran lo más representativos posible del grupo etario. Luego de este segundo filtro, que tuvo el objetivo de evitar edades extremas, se seleccionaron aleatoriamente a los niños que serían evaluados.

Procedimiento. Previo a la recogida de datos, se realizó una capacitación en cada una de las universidades socias del proyecto. Para esto, se envió a un integrante del equipo central en Santiago a cada una de las localidades para realizar una capacitación práctica de ocho horas en la aplicación del instrumento. Es importante notar que las personas que fueron seleccionadas para realizar las aplicaciones eran psicólogos con experiencia previa en el uso de instrumentos similares a WISC-V. Posterior a la capacitación, se pidió a los evaluadores que enviaran su primer protocolo para revisión. Esta revisión tuvo el objetivo de verificar que los procedimientos de aplicación y corrección fueran realizados correctamente. Se retroalimentó a los

evaluadores que habían cometido algún error de aplicación o corrección y se les pidió que enviaran un segundo protocolo para revisión. Los evaluadores que finalmente participaron en el proyecto fueron aquellos cuyos protocolos no tenían errores de aplicación. Es importante notar que, dadas las características de este instrumento, todas las evaluaciones se realizaron de forma individual y duraron entre 60 a 90 minutos, aunque esto varió en función de la edad.

Tabla 3
Composición de la Muestra de Estandarización

Región	Municipal		Particular subvencionado		Particular pagado		Total
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	
Arica y Parinacota	10	10	12	9	12	12	65
Antofagasta	10	9	7	11	12	12	61
Metropolitana	51	61	52	62	43	38	307
Maule	18	19	10	11	10	12	80
Biobío	21	13	15	22	18	17	106
Araucanía	25	21	9	11	10	11	87
Los Lagos	9	9	5	9	8	8	48
Subtotal	144	142	110	135	113	110	
Total	286		245		223		754

Una vez recogidos los datos, la información de cada participante fue subida a una base de datos en línea para tener un registro del avance del proyecto. Los protocolos fueron corregidos por los evaluadores de cada región, quienes posteriormente enviaron el documento original para ser nuevamente corregido por el equipo central del proyecto en Santiago, con el objetivo de asegurar la calidad de los datos. Finalmente, los datos fueron digitados en una base en la que se omitió la información personal de cada evaluado.

Construcción de Normas. Para la construcción de la norma nacional se utilizó el método de norma inferencial (Zhu & Chen, 2011), tanto para las subpruebas como para los puntajes compuestos. Este proceso se utiliza para obtener normas robustas usando un menor número de casos que con procedimientos tradicionales de creación de normas. Las normas inferenciales toman como punto de partida los datos obtenidos en la muestra, los que luego son ajustados de acuerdo con criterios generales de progresión (e.g., al considerar la progresión de cada subprueba en función de la edad) como también con los antecedentes del desarrollo. Para más información sobre este procedimiento, se sugiere revisar Zhu y Chen (2011).

Para obtener los puntajes estandarizados de las subpruebas se identificaron diversos momentos de cada puntaje, considerando las medias, desviación estándar y asimetría de la distribución. Posteriormente, considerando todos los momentos, se trazaron distintas curvas de progresión, utilizando diversas regresiones polinomiales para identificar aquellas curvas que mejor se ajustaban al desarrollo cognitivo en función de los antecedentes teóricos, en conjunto con el patrón de desarrollo observado en WISC-V. Fueron estas curvas las que se utilizaron para generar las distribuciones de puntajes brutos para cada grupo etario. Luego, estos puntajes fueron transformados a puntajes escalares. Por último, se analizó la progresión de puntajes a nivel intra e inter-grupo etario y se corrigieron pequeñas irregularidades, a través de un proceso de suavizado para obtener un continuo de puntajes que fuese coherente con las teorías del desarrollo.

Para construir los puntajes compuestos, primero se sumaron los puntajes escalares de las subpruebas que los componen (e.g., los puntajes escalares de AN y VOC se sumaron para obtener la suma de puntajes escalares de ICV). Después, se transformó la suma de puntajes escala a una distribución normal, utilizando la métrica estándar (Media = 100 y $DE = 15$). Por último, se corrigieron pequeñas irregularidades a través de un proceso de suavizado lineal para asegurar una distribución normal de varios intervalos de puntaje manteniendo los puntajes compuestos lo más cerca posible a una media de 100 y una desviación estándar de 15.

Resultados. En la Tabla 4 se muestran los resultados del proceso de construcción de datos normativos, utilizando el procedimiento de norma inferencial a nivel de subpruebas. Se puede observar que para todas las subpruebas se mantiene una media de puntaje de 10 y una desviación estándar de 3. Es importante notar que la máxima desviación posible a estos puntajes es de 0,1 para asegurar que las subpruebas cumplan con la métrica escalar.

Tabla 4
Puntajes Escalares de Subpruebas Utilizando Procedimiento de Norma Inferencial

Subprueba	Sigla	Media	DE
Construcción con Cubos	CC	10,02	3,01
Analogías	AN	10,00	3,00
Matrices de Razonamiento	MR	10,01	3,03
Retención de Dígitos	RD	10,05	3,02
Claves	CLA	10,09	2,95
Vocabulario	VOC	10,00	3,05
Balanzas	BAL	10,02	3,05
Rompecabezas Visuales	RV	10,00	3,02
Retención de Imágenes	RI	10,06	3,06
Búsqueda de Símbolos	BS	10,00	3,07
Información	INF	10,03	3,06
Secuenciación de Letras - Números	SLN	9,98	3,04
Cancelación	CAN	10,07	3,02
Comprensión	COM	10,01	3,01
Aritmética	ARI	10,01	3,00

En la Tabla 5 se muestran los resultados de la aplicación del procedimiento de norma inferencial a los puntajes compuestos de los índices. Se puede observar que todos los puntajes tienen una media de 100 y una desviación estándar de 15. Es importante notar que para estos puntajes se admite una desviación de 0,5 máximo para asegurar que se mantenga la métrica estándar.

Tabla 5
Puntajes Estándar de Índices Utilizando el Procedimiento de Norma Inferencial

Índice	Sigla	Media	DE
Coficiente Intelectual Total	CIT	100,01	15,01
Índice de Comprensión Verbal	ICV	100,00	15,01
Índice de Razonamiento Fluido	IRF	100,04	15,02
Índice Visoespacial	IVE	100,01	15,05
Índice de Memoria de Trabajo	IMT	99,67	15,05
Índice de Velocidad de Procesamiento	IVP	100,07	15,01
Índice de Memoria de Trabajo Auditiva	IMTA	100,00	15,08
Índice de Razonamiento Cuantitativo	IRC	100,25	15,07
Índice de Competencia Cognitiva	ICC	99,97	15,08
Índice de Habilidad General	IHG	100,07	15,04
Índice No Verbal	INV	100,03	15,04

En la Tabla 6 se presenta una comparación entre los porcentajes esperados de acuerdo a una distribución normal y los porcentajes obtenidos en la muestra para varios intervalos de puntaje de CIT. Al realizar una prueba χ^2 de bondad de ajuste, se observa que la distribución observada en la muestra no presenta diferencias estadísticamente significativas con la distribución de una curva normal ($\chi^2(6) = 2,97, p = 0,813$).

Tabla 6
Comparación de Distribución de Puntajes de CIT entre la Muestra y la Curva Normal

Rango de puntaje	Curva normal	Muestra
130-160	2,49%	1,79%
120-129	7,22%	6,61%
110-119	16,65%	18,18%
90-109	49,48%	49,86%
80-89	15,61%	15,43%
70-79	6,49%	6,06%
40-69	2,10%	2,07%

Dado que los resultados de la aplicación del procedimiento de norma inferencial cumplen con los requisitos psicométricos, tanto para las subpruebas como para los índices, se utilizaron dichos resultados como datos normativos.

Evidencia de Confiabilidad

Análisis de Datos

Para obtener evidencia de confiabilidad de las subpruebas se utilizó el procedimiento de split-half para todas las pruebas de la batería, a excepción de CLA, BS y CAN. Los coeficientes de confiabilidad a nivel de subprueba fueron obtenidos a través de la correlación directa entre las dos mitades de la prueba corregidos por fórmula de Spearman-Brown para el total de la subprueba (Crocker & Algina, 1986; Haertel, 2006; Li et al., 1996). Para las subpruebas de CLA, BS y CAN se utilizó el procedimiento de medidas repetidas para obtener evidencia de confiabilidad, dado que las características de estas subpruebas hacen que el procedimiento de split-half no sea adecuado.

Para obtener la consistencia interna de los puntajes compuestos se calcularon los coeficientes de confiabilidad con los mismos procedimientos que en la versión norteamericana, utilizando la fórmula propuesta por Guilford (1954), Haertel (2006) y Nunnally y Bernstein (1994). Los coeficientes de confiabilidad promedio se calcularon utilizando la transformación z de Fisher (Silver & Dunlap, 1987; Strube, 1988).

Resultados

En la Tabla 7 se muestran los resultados de confiabilidad para todos los indicadores de WISC-V. Los resultados indican que los coeficientes de confiabilidad a nivel de las subpruebas van desde un rango bueno, en CAN con 0,645, hasta un rango excelente, en Balanzas (BAL) con 0,941, utilizando los descriptores de rango indicados por Cicchetti (1994). De acuerdo con este mismo autor, los coeficientes de confiabilidad de todos los índices se encuentran en rango excelente.

Evidencia de Validez

Se presentan los resultados obtenidos por distintas muestras, utilizando los puntajes estandarizados con el procedimiento de norma inferencial.

Estructura Interna

Análisis de Datos. Para dar cuenta de la estructura interna de la prueba se realizaron dos análisis: primero, la correlación directa entre subpruebas e índices y segundo, un análisis factorial confirmatorio (AFC).

Tabla 7
Coefficientes de Confiabilidad de Subpruebas e Índices de WISC-V

Subprueba/índice	Grupo etario											Promedio
	6	7	8	9	10	11	12	13	14	15	16	
CC	0,806	0,811	0,807	0,837	0,840	0,830	0,843	0,794	0,797	0,853	0,838	0,824
AN	0,885	0,881	0,909	0,939	0,942	0,936	0,926	0,926	0,920	0,931	0,916	0,921
MR	0,885	0,893	0,916	0,922	0,907	0,907	0,873	0,893	0,893	0,910	0,896	0,900
RD	0,891	0,911	0,909	0,914	0,914	0,902	0,895	0,897	0,922	0,933	0,871	0,907
CLA	0,778	0,880	0,859	0,790	0,944	0,618	0,925	0,891	0,911	0,954	0,981	0,898
VOC	0,701	0,766	0,881	0,914	0,896	0,889	0,926	0,910	0,904	0,930	0,909	0,888
BAL	0,907	0,902	0,938	0,942	0,940	0,953	0,948	0,937	0,948	0,951	0,960	0,941
RV	0,876	0,888	0,891	0,899	0,908	0,915	0,917	0,909	0,912	0,923	0,885	0,903
RI	0,855	0,893	0,897	0,909	0,886	0,893	0,899	0,907	0,905	0,901	0,830	0,891
BS	0,846	0,833	0,677	0,603	0,917	0,429	0,791	0,884	0,924	0,814	0,920	0,822
INF	0,797	0,863	0,897	0,922	0,929	0,926	0,904	0,917	0,933	0,938	0,921	0,910
SLN	0,935	0,903	0,886	0,902	0,928	0,886	0,853	0,860	0,855	0,912	0,891	0,895
CAN	0,551	0,663	0,465	0,501	0,784	0,733	0,685	0,663	0,694	0,623	0,636	0,645
COM	0,686	0,773	0,881	0,831	0,900	0,897	0,926	0,920	0,902	0,915	0,861	0,876
ARI	0,844	0,878	0,883	0,889	0,919	0,911	0,895	0,915	0,924	0,922	0,897	0,900
CCp	0,845	0,846	0,843	0,874	0,854	0,856	0,855	0,775	0,810	0,845	0,798	0,838
CCsb	0,806	0,810	0,808	0,834	0,838	0,829	0,847	0,838	0,810	0,859	0,847	0,830
RDd	0,666	0,751	0,741	0,769	0,742	0,730	0,717	0,773	0,780	0,799	0,823	0,757
RDi	0,853	0,847	0,838	0,877	0,838	0,775	0,783	0,837	0,855	0,850	0,585	0,822
RDs	0,804	0,825	0,834	0,837	0,814	0,838	0,813	0,756	0,810	0,899	0,814	0,825
CAa	0,607	0,502	0,735	0,624	0,709	0,504	0,773	0,426	0,671	0,862	0,662	0,662
CAe	0,351	0,868	0,885	0,564	0,828	0,429	0,531	0,612	0,926	0,983	0,912	0,804
ICV	0,874	0,897	0,935	0,955	0,953	0,952	0,957	0,953	0,946	0,961	0,948	0,943
IRF	0,915	0,906	0,952	0,954	0,943	0,955	0,940	0,938	0,953	0,959	0,961	0,945
IVE	0,889	0,900	0,907	0,918	0,912	0,923	0,921	0,906	0,909	0,925	0,913	0,912
MT	0,900	0,930	0,942	0,941	0,937	0,931	0,936	0,938	0,949	0,950	0,883	0,933
IVP	0,857	0,904	0,858	0,813	0,954	0,672	0,895	0,910	0,946	0,921	0,959	0,900
IMTA	0,946	0,944	0,939	0,944	0,952	0,932	0,924	0,928	0,934	0,953	0,928	0,939
IRC	0,885	0,909	0,941	0,947	0,949	0,957	0,942	0,952	0,959	0,961	0,955	0,945
ICC	0,912	0,939	0,940	0,923	0,963	0,851	0,936	0,942	0,963	0,958	0,930	0,937
IHG	0,921	0,928	0,961	0,969	0,966	0,969	0,965	0,960	0,964	0,972	0,968	0,961
INV	0,937	0,950	0,968	0,963	0,966	0,955	0,960	0,956	0,966	0,975	0,965	0,961
CIT	0,937	0,947	0,971	0,971	0,974	0,966	0,969	0,965	0,973	0,982	0,973	0,968

El desarrollo de la versión norteamericana de WISC-V estuvo guiado por los supuestos teóricos de que la escala evalúa la habilidad cognitiva general a través de cinco dominios: ICV, IVE, IRF, IMT e IVP. Esta estructura está organizada en un modelo jerárquico de dos niveles. El primero incluye los cinco factores de primer orden, correspondientes a los dominios descritos previamente; el segundo nivel corresponde al de inteligencia general. Esta estructura fue puesta a prueba con un AFC con el que se evaluó su ajuste. Los resultados mostraron que la estructura propuesta tenía excelentes grados de ajuste, incluso mejores que los modelos en los que se proponía una estructura de cuatro factores (Wechsler, 2014).

El desarrollo de la versión chilena tomó como punto de partida el modelo propuesto en la versión norteamericana. En consecuencia, se realizó un AFC, utilizando el programa MPlus 7.0 para poner a prueba la misma estructura de la versión original.

Posteriormente, se puso a prueba el mismo modelo de cinco dominios, pero esta vez en función de las 10 subpruebas primarias. Para ello, se utilizaron los índices de ajuste χ^2/gl , TLI, CFI y RMSEA. Se siguieron las recomendaciones de varios autores (Pérez et al., 2013; Sun, 2005) para identificar las características de un buen ajuste, a saber, una proporción de χ^2/gl menor a 3, TLI y CFI por sobre 0,95 y un RMSEA menor a 0,06.

Resultados. La Tabla 8 contiene la matriz de correlaciones entre subpruebas e índices. Se puede observar ahí que las subpruebas de cada dominio son aquellas que generalmente tienen una mayor correlación entre sí. Esto se observa particularmente en ICV, en el que VOC y AN tienen una correlación de 0,664, y también entre AN e INF, cuya correlación es de 0,637.

A nivel de subpruebas e índices, las mayores correlaciones se observan entre VOC e ICV (0,914), CC e IVE (0,875), BAL e IRF (0,851), RI e IMT (0,877) y BS e IVP (0,862).

Tabla 8
Matriz de Correlaciones de Subpruebas e Índices

Subprueba/ índice	Dominio de comprensión verbal			Dominio visoespacial		Dominio de razonamiento fluido			Dominio de memoria de trabajo			Dominio de velocidad de procesamiento					Puntajes compuestos								
	AN	VOC	INF	COM	CC	RV	MR	BAL	ARI	RD	RI	SLN	CLA	BS	CAN	ICV	IVE	IRF	IMT	IVP	IRC	IMTA	INV	IHG	ICC
AN	1																								
VOC	0,664	1																							
INF	0,637	0,624	1																						
COM	0,539	0,599	0,510	1																					
CC	0,414	0,359	0,347	0,269	1																				
RV	0,465	0,439	0,390	0,291	0,531	1																			
MR	0,435	0,431	0,365	0,337	0,421	0,519	1																		
BAL	0,390	0,393	0,416	0,329	0,370	0,412	0,442	1																	
ARI	0,519	0,515	0,543	0,415	0,413	0,411	0,456	0,459	1																
RD	0,482	0,450	0,503	0,357	0,364	0,389	0,442	0,413	0,617	1															
RI	0,406	0,410	0,401	0,316	0,333	0,408	0,420	0,339	0,445	0,534	1														
SLN	0,457	0,430	0,498	0,389	0,367	0,373	0,397	0,373	0,595	0,635	0,481	1													
CLA	0,313	0,325	0,321	0,312	0,304	0,283	0,282	0,220	0,355	0,368	0,283	0,355	1												
BS	0,287	0,300	0,295	0,263	0,288	0,304	0,284	0,208	0,288	0,298	0,265	0,321	0,466	1											
CAN	0,210	0,230	0,240	0,175	0,212	0,178	0,144	0,141	0,183	0,164	0,143	0,173	0,318	0,420	1										
ICV	0,909	0,914	0,695	0,623	0,420	0,494	0,470	0,428	0,572	0,510	0,448	0,487	0,350	0,324	0,247	1									
IVE	0,500	0,455	0,422	0,320	0,875	0,874	0,537	0,445	0,472	0,430	0,426	0,425	0,339	0,343	0,230	0,520	1								
IRF	0,485	0,485	0,460	0,393	0,464	0,549	0,846	0,851	0,540	0,505	0,447	0,454	0,293	0,288	0,168	0,529	0,578	1							
IMT	0,509	0,491	0,515	0,384	0,399	0,457	0,493	0,430	0,606	0,874	0,877	0,636	0,370	0,320	0,175	0,548	0,491	0,544	1						
IVP	0,347	0,365	0,358	0,335	0,345	0,344	0,330	0,250	0,375	0,387	0,319	0,393	0,850	0,862	0,433	0,392	0,398	0,340	0,402	1					
IRC	0,531	0,533	0,561	0,436	0,456	0,481	0,523	0,856	0,851	0,602	0,460	0,566	0,336	0,288	0,190	0,586	0,535	0,815	0,607	0,364	1				
IMTA	0,524	0,493	0,559	0,417	0,409	0,427	0,468	0,439	0,674	0,903	0,563	0,903	0,408	0,348	0,194	0,557	0,480	0,535	0,835	0,439	0,650	1			
INV	0,583	0,570	0,539	0,447	0,714	0,763	0,745	0,676	0,611	0,604	0,675	0,565	0,565	0,433	0,268	0,630	0,843	0,836	0,731	0,581	0,752	0,652	1		
IHG	0,779	0,774	0,652	0,561	0,692	0,638	0,736	0,703	0,643	0,583	0,518	0,552	0,390	0,369	0,257	0,851	0,759	0,848	0,630	0,443	0,787	0,634	0,889	1	
ICC	0,512	0,512	0,521	0,430	0,443	0,474	0,492	0,404	0,585	0,757	0,722	0,615	0,722	0,698	0,353	0,562	0,527	0,527	0,843	0,828	0,578	0,762	0,783	0,641	1
CIT	0,757	0,750	0,665	0,561	0,667	0,621	0,710	0,666	0,688	0,722	0,559	0,618	0,569	0,434	0,292	0,827	0,736	0,811	0,732	0,585	0,791	0,747	0,916	0,963	0,787

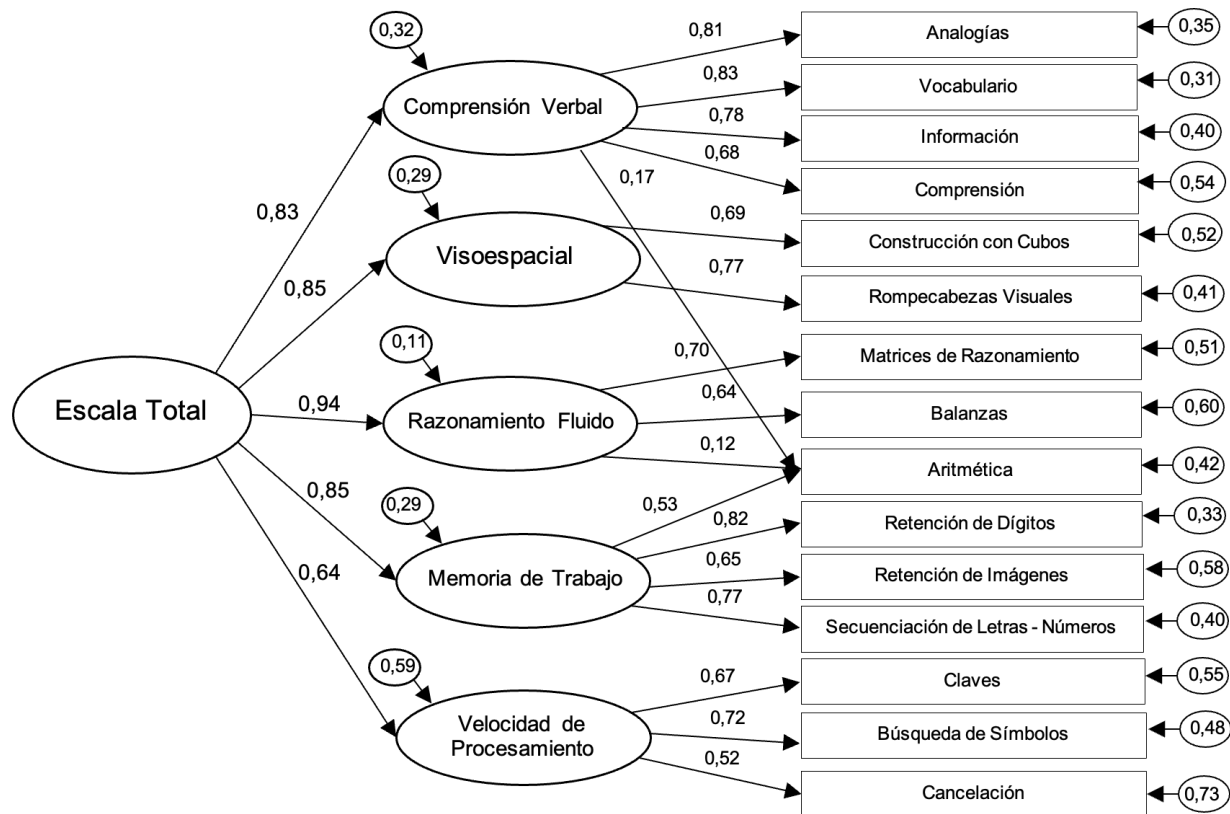
Nota. Todas las correlaciones son significativas al nivel de $p < 0,001$.

Análisis Factorial Confirmatorio

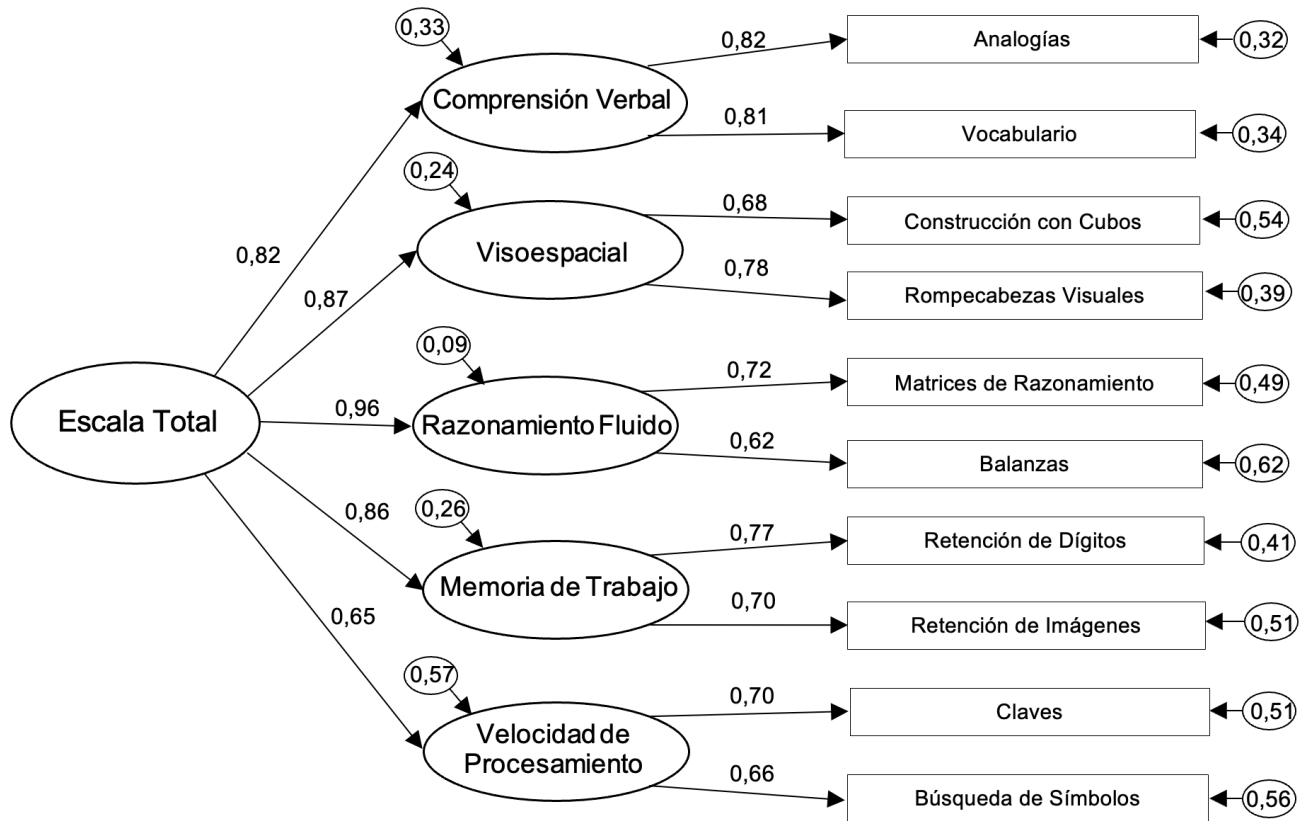
Los resultados del AFC mostraron que el modelo de cinco factores logra excelentes grados de ajuste para la muestra chilena, $\chi^2(82, n = 754) = 189,22, p < 0,001, \chi^2/gl = 2,3, RMSEA = 0,042, CFI = 0,977, TLI = 0,97$. Estos resultados permiten replicar el modelo general de la batería original norteamericana. En la Figura 2 se presentan las cargas factoriales, correlaciones y residuos del modelo, considerando las 15 subpruebas del instrumento, en función de los cinco dominios como factor de primer orden e inteligencia general como factor de segundo orden.

Figura 2

Modelo Jerárquico de Cinco Factores para las Subpruebas Primarias y Complementarias



Al poner a prueba el mismo modelo de cinco dominios en función de las 10 subpruebas primarias, los resultados nuevamente mostraron un excelente grado de ajuste, $\chi^2(30, n = 754) = 52,245, p < 0,001, \chi^2/gl = 1,7, RMSEA = 0,031, CFI = 0,99, TLI = 0,99$. En la Figura 3 se presentan las cargas factoriales, correlaciones y residuos del modelo, considerando las 10 subpruebas primarias.

Figura 3*Modelo Jerárquico de Cinco Factores para las 10 Subpruebas Primarias*

Muestras Especiales

Se aplicó la batería a cinco muestras especiales: discapacidad intelectual, ruralidad, trastorno del espectro autista, trastorno por déficit atencional con hiperactividad (TDAH) y TEL.

Participantes

Todos los participantes de las muestras especiales fueron parte de distintas muestras intencionadas que fueron evaluadas por distintos equipos. Las muestras de discapacidad intelectual ($n = 24$), trastorno del espectro autista ($n = 26$) y TDAH ($n = 27$) fueron recogidas en el Hospital San Borja Arriarán en Santiago por profesionales de dicho centro. Las muestras rurales ($n = 47$) fueron recogidas en la zona central y sur de Chile, dado que son las zonas en las que hay un mayor porcentaje de población rural, por la Universidad Católica del Maule en Talca, que recogió muestra en la localidad del Maule, y la Universidad de La Frontera en Temuco, que recogió muestra en la localidad de la araucanía. La muestra de sujetos con TEL ($n = 56$) fue recogida por el equipo de Santiago en colegios municipales de la comuna de San Bernardo.

La selección de los sujetos de los grupos de discapacidad intelectual, trastorno del espectro autista, TDAH y TEL se realizó considerando evaluaciones previas que confirmaran los diagnósticos de estos grupos. Todos los evaluados del grupo con trastorno del espectro autista habían sido evaluados previamente con la Escala de Observación para el Diagnóstico de Autismo – 2 (Lord et al., 2012). Los evaluados del grupo con TEL tenían evaluaciones previas de fonoaudiólogos y psicólogos. Para seleccionar los sujetos del grupo de muestra rural se siguieron los lineamientos de investigaciones anteriores que han abordado la temática rural utilizando las escalas Wechsler en Chile (Fuica et al., 2014).

Resultados

Los resultados de la aplicación a las muestras especiales se muestran en la Tabla 9. En general, se observa que en la mayoría de las subpruebas y los índices las muestras especiales obtienen rendimientos menores que la muestra normativa. Lo anterior es coherente con lo esperado para los diagnósticos y grupos que componen esta muestra. Es esperable que personas con el diagnóstico de discapacidad intelectual obtengan menores resultados en pruebas que evalúen la cognición, lo que se observa en todos los índices que se incluyen en WISC-V.

Tabla 9
Resultados de WISC-V en Muestras Especiales

Subprueba/ índice	Discapacidad intelectual (n = 24)		Ruralidad (n = 47)		Trastorno del espectro autista (n = 26)		Trastorno por déficit atencional con hiperactividad (n = 27)		Trastorno específico del lenguaje (n = 56)	
	Media	DE	Media	DE	Media	DE	Media	DE	Media	DE
CC	4,7	2,9	9,5	2,6	10,9	3,9	9,3	2,9	8,5	2,7
AN	4,4	1,6	8,9	2,5	9,4	2,4	8,7	2,3	6,9	3,2
MR	3,4	2,1	9,5	2,7	9,9	3,2	8,8	3,3	7,0	3,5
RD	3,0	2,2	9,9	2,3	8,3	3,6	8,0	2,1	6,9	2,7
CLA	3,5	3,0	10,0	2,4	5,8	3,3	7,9	3,9	7,6	2,6
VOC	3,2	1,7	9,1	2,7	8,3	3,1	8,2	2,5	7,0	2,2
BAL	5,1	2,7	10,2	2,8	10,4	2,9	9,2	3,2	8,4	2,9
RV	4,4	2,6	9,6	2,4	10,0	3,1	8,8	3,0	8,1	2,6
RI	3,6	2,3	9,3	2,4	8,9	3,2	8,8	3,4	7,7	2,6
BS	3,6	2,4	9,3	2,3	6,1	2,2	7,6	2,8	7,6	3,1
INF	3,7	1,9	8,6	2,5	8,3	3,3	8,0	2,4	6,1	2,1
SLN	1,9	1,6	9,5	1,9	8,8	4,7	8,2	2,8	6,2	2,1
CAN	4,5	3,1	9,6	3,3	5,8	3,4	8,7	4,2	8,6	2,6
COM	4,2	2,1	8,9	2,2	6,8	3,0	8,1	2,4	7,3	2,4
ARI	2,5	1,8	9,9	2,4	8,3	3,1	8,3	2,6	7,2	2,3
ICV	65,6	9,5	94,6	12,1	94,0	10,7	91,7	11,6	83,1	13,1
IVE	68,1	14,9	97,4	12,0	102,2	18,7	94,6	13,6	90,5	12,7
IRF	67,3	12,8	99,2	13,5	101,0	15,6	94,2	16,4	86,8	12,7
IMT	61,8	11,2	97,3	10,8	91,4	16,6	90,5	12,5	84,6	11,5
IVP	61,6	15,5	97,7	11,4	76,5	14,6	86,7	17,0	85,5	12,7
IRC	64,5	10,8	100,6	13,1	96,4	14,5	92,7	14,3	87,4	11,9
IMTA	56,7	10,9	98,5	9,7	92,3	18,7	90,8	11,6	82,0	10,4
INV	59,1	12,3	97,4	11,4	94,7	17,6	90,8	12,7	84,1	12,1
IHG	60,3	9,7	96,2	12,9	99,9	14,0	92,3	13,0	83,8	10,8
ICC	56,3	11,3	97,0	10,9	81,1	16,2	86,6	12,9	82,3	11,4
CIT	57,7	9,8	96,7	12,3	93,7	15,0	89,4	10,3	81,7	9,9

En la muestra rural también se observan menores desempeños en la mayoría de los índices, lo que resulta esperable en este grupo, dadas las menores oportunidades educativas disponibles en contextos rurales. Para más información sobre los resultados de esta muestra se recomienda revisar Rodríguez et al. (2019).

Los resultados de la muestra que incluye a personas con trastorno del espectro autista también presentan menores desempeños que los de la muestra normativa en la mayoría de los índices, a excepción de IVE e IRF. Es importante notar que los sujetos que fueron incluidos en este grupo no presentaban un compromiso en el lenguaje.

El grupo de sujetos con el diagnóstico de TDAH también presenta menores desempeños que la muestra normativa. Es importante notar que las personas evaluadas en este grupo son de NSE bajo. Dado que en Chile existe evidencia que ha mostrado la relación entre desempeño cognitivo y NSE (Rosas & Santa Cruz, 2013), es posible que en los resultados de esta muestra puedan estarse confundiendo los efectos del

diagnóstico con los de las oportunidades educativas asociadas a un menor NSE. Es necesario realizar más investigaciones sobre este diagnóstico en el contexto nacional para obtener información adicional que permita complementar los resultados de estos datos.

Por último, los resultados del grupo con TEL aparecen particularmente relevantes, dado que no es posible identificar un menor rendimiento específicamente en los componentes asociados al lenguaje, como podrían ser aquellos que están siendo evaluados por el ICV. Si bien los sujetos de este grupo también pertenecen a un NSE bajo y, por tanto, es necesario considerar las mismas restricciones a la interpretación de resultados que para el grupo con TDAH, es importante notar que en el grupo con TEL se observan rendimientos similares en el ICV a los de otros índices. La única excepción a lo anterior se observa en el IVE; pese a ello, todos los índices restantes se encuentran en el mismo rango que el ICV. Esto hace necesario revisar con más atención la evaluación de este diagnóstico, puesto que no parece observarse un menor desempeño específico en los elementos asociados al lenguaje, sino uno menor en la mayoría de los indicadores cognitivos.

Correlación con Otras Mediciones

Participantes

Para evaluar la validez convergente de WISC-V, se aplicaron las baterías WISC-III y WAIS-IV de forma contrabalanceada a una muestra por conveniencia de 121 sujetos (81 sujetos WISC-III; 40 sujetos WAIS-IV) provenientes de Arica, Santiago, Talca, Temuco y Puerto Montt. Estos sujetos no formaron parte de la muestra de tipificación de la batería.

Procedimiento

Se aplicaron las 13 subpruebas necesarias para obtener los subíndices de WISC-III. En WAIS-IV, se aplicaron las 10 subpruebas centrales para obtener los índices y el CIT. Por último, en WISC-V, se aplicaron las 10 subpruebas necesarias para obtener los índices principales y el CIT.

Análisis de Datos

Para obtener evidencia de validez convergente, se realizaron análisis de correlaciones directas entre las subpruebas e índices similares para las comparaciones entre WISC-V y WAIS-IV, y entre WISC-V y WISC-IIIv.ch.

Resultados

La comparación de resultados entre WISC-V y WISC-III se encuentra en la Tabla 10. Las correlaciones entre todas las comparaciones se ubican entre 0,645 para BS hasta 0,899 en el indicador general de CIT.

La Tabla 11 muestra los resultados de las comparaciones entre WISC-V y WAIS-IV. A nivel de correlaciones, todas las comparaciones entre índices y subpruebas van desde 0,447 entre IRF y RP hasta 0,838 en ICV.

Resultados por NSE y Sexo

Análisis de Datos

Para el análisis de datos de la comparación entre los resultados por NSE se utilizó un ANOVA dado que los grupos de interés son más de dos. Para el análisis de datos en la comparación de resultados por género, también se utilizó un ANOVA, aunque se utilizó una prueba t para muestras independientes con el objetivo de obtener los intervalos de confianza para la comparación de medias.

Tabla 10*Correlaciones a Nivel de Índices y Subpruebas entre WISC-V y WISC-IIIv.ch*

Par de índices/subprueba	Sigla	Correlación
Coficiente Intelectual Total	CIT	0,899
Índice de Comprensión Verbal	ICV	0,864
Índice Visoespacial - Organización Perceptual	IVE - OP	0,792
Índice de Razonamiento Fluido - Organización Perceptual	IRF - OP	0,677
Índice Memoria de Trabajo - Ausencia de Distractibilidad	IMT - AD	0,773
Índice de Velocidad de Procesamiento	IVP	0,780
Construcción con Cubos	CC	0,749
Analogías	AN	0,737
Claves	CLA	0,782
Vocabulario	VOC	0,712
Retención de Dígitos	RD	0,771
Búsqueda de Símbolos	BS	0,645

Nota. Todas las correlaciones son significativas al nivel de $p < 0,001$.

Tabla 11*Correlaciones a Nivel de Subprueba e Índices entre WISC-V y WAIS-IV*

Par de índices/subprueba	Sigla	Correlación
Coficiente Intelectual Total	CIT	0,825
Índice de Comprensión Verbal	ICV	0,838
Índice Visoespacial - Razonamiento Perceptual	IVE - RP	0,627
Índice de Razonamiento Fluido - Razonamiento Perceptual	IRF - RP	0,447**
Índice Memoria de Trabajo	IMT	0,689
Índice de Velocidad de Procesamiento	IVP	0,785
Construcción con Cubos	CC	0,552
Analogías	AN	0,554
Matrices de Razonamiento	MR	0,533
Retención de Dígitos	RD	0,748
Claves	CLA	0,696
Vocabulario	VOC	0,789
Rompecabezas Visuales	RV	0,626
Búsqueda de Símbolos	BS	0,635

Nota. Todas las correlaciones son significativas al nivel de $p < 0,001$, con excepción de **, que se encuentra al nivel de $p < 0,05$.

Resultados

En la Tabla 12 se muestran los resultados que obtienen en promedio en CIT y en los índices principales los niños de las tres dependencias educativas que se consideraron como parte del muestreo para la estandarización. Se puede observar que el grupo de participantes pertenecientes a establecimientos educativos de dependencia particular obtiene mejores resultados en todos los indicadores de la batería, seguidos luego de los particulares subvencionados y finalmente los municipales.

No se observó homogeneidad de varianzas en el indicador de CIT, por lo que se utilizó la corrección Welch para interpretar el ANOVA. Los resultados en este indicador dan cuenta de una diferencia estadísticamente significativa entre los grupos ($F(2, 477) = 89,03, p < 0,001$) con un tamaño de efecto calculado por eta cuadrado que se considera grande ($\eta^2 = 0,191$). Al realizar el análisis post hoc usando la prueba de Tukey, se observa que todas las comparaciones son significativas al nivel de $p < 0,001$; la diferencia media entre el nivel particular subvencionado y municipal fue de 4,766 (IC 95% [1,937, 7,594]). El tamaño de efecto de esta comparación es medio, dado que se observa una d de Cohen de 0,342. La diferencia entre el particular pagado

y el particular subvencionado fue de 11,028 (IC 95% [8,022, 14,034]); en este caso, el tamaño de efecto es grande dado que la d de Cohen es de 0,859. Por último, la diferencia media entre el particular pagado y el municipal fue de 15,794 (IC 95% [12,919, 18,668]); al igual que en la anterior, el tamaño de efecto de esta comparación es grande al tomar en cuenta que la d de Cohen es de 1,153.

Tabla 12*Comparación de Resultados de WISC-V por Dependencia Escolar*

Dependencia		CIT	ICV	IRF	IVE	IMT	IVP
Municipal	Media	93,8	93,4	95,3	97,2	94,9	96,3
	DE	14,6	14,5	15,3	15,8	15,2	14,4
Particular subvencionado	Media	98,5	99,0	99,5	97,5	99,3	99,5
	DE	13,1	12,7	14,3	13,5	13,3	14,6
Privado	Media	109,6	109,5	106,7	106,6	106,3	105,7
	DE	12,5	13,0	12,8	13,6	14,5	14,7

Se observó homogeneidad de las varianzas en el indicador de ICV. El análisis ANOVA permitió identificar diferencias significativas entre las tres dependencias ($F(2, 727) = 87,99, p < 0,001$) con un tamaño de efecto grande ($\eta^2 = 0,195$). Al realizar el análisis post hoc usando la prueba de Tukey, se observa que todas las comparaciones son significativas el nivel de $p < 0,001$; la diferencia media entre el nivel particular subvencionado y municipal fue de 5,527 (IC 95% [2,716, 8,338]). El tamaño de efecto de esta comparación es medio, dado que se observa una d de Cohen de 0,403. La diferencia entre el particular pagado y el particular subvencionado fue de 10,539 (IC 95% [7,552, 13,526]); en este caso, el tamaño de efecto es grande dado que la d de Cohen es de 0,821. Por último, la diferencia media entre el particular pagado y el municipal fue de 16,066 (IC 95% [13,206, 18,927]); al igual que en la anterior, el tamaño de efecto de esta comparación es grande al tomar en cuenta que la d de Cohen es de 1,162.

No se observó homogeneidad de varianzas en el indicador de IRF, por lo que se utilizó la corrección Welch para interpretar el ANOVA. Los resultados en este indicador dan cuenta de una diferencia estadísticamente significativa entre los grupos ($F(2, 494) = 42,70, p < 0,001$) con un tamaño de efecto calculado por eta cuadrado que se considera entre mediano y grande ($\eta^2 = 0,096$). Al realizar el análisis post hoc usando la prueba de Tukey, se observa que todas las comparaciones son significativas el nivel de $p < 0,001$ con excepción de la comparación entre el nivel municipal y el particular subvencionado, que es significativa al nivel de 0,002. La diferencia media entre el nivel particular subvencionado y municipal fue de 4,192 (IC 95% [1,269, 7,115]). El tamaño de efecto de esta comparación es medio, dado que se observa una d de Cohen de 0,282. La diferencia entre el particular pagado y el particular subvencionado fue de 7,199 (IC 95% [4,078, 10,321]); en este caso, el tamaño de efecto es también medio dado que la d de Cohen es de 0,528. Por último, la diferencia media entre el particular pagado y el municipal fue de 11,391 (IC 95% [8,383, 14,400]); al igual que en la anterior, el tamaño de efecto de esta comparación es medio al tomar en cuenta que la d de Cohen es de 0,798.

Se observó homogeneidad de las varianzas en el indicador de IVE. El análisis ANOVA permitió identificar diferencias significativas entre las dependencias IVE: ($F(2, 748) = 32,74, p < 0,001$) con un tamaño de efecto medio ($\eta^2 = 0,08$). Al realizar el análisis post hoc usando la prueba de Tukey, se observa que la diferencia entre las medias de la dependencia municipal y particular subvencionada no es significativa. Solo son significativas las diferencias entre la dependencia municipal y particular pagada así como también la diferencia entre el particular subvencionado y el particular pagado, ambas a un nivel de $p < 0,001$. La diferencia entre el particular pagado y el particular subvencionado fue de 9,136 (IC 95% [5,981, 12,291]); en este caso, el tamaño de efecto es medio dado que la d de Cohen es de 0,673. La diferencia media entre el particular pagado y el municipal fue de 9,563 (IC 95% [6,522, 12,604]); al igual que en la anterior, el tamaño de efecto de esta comparación es medio al tomar en cuenta que la d de Cohen es de 0,643.

Se observó homogeneidad de las varianzas en el indicador de IMT. El análisis ANOVA permitió identificar diferencias significativas entre las tres dependencias ($F(2, 748) = 39,05, p < 0,001$) con un tamaño de efecto entre mediano y grande ($\eta^2 = 0,095$). Al realizar el análisis post hoc usando la prueba de Tukey, se observa que todas las comparaciones son significativas el nivel de $p < 0,001$, con excepción de la comparación

entre el nivel municipal y el particular subvencionado, que es significativa al nivel de 0,001. La diferencia media entre el nivel particular subvencionado y municipal fue de 4,357 (IC 95% [1,426, 7,289]). El tamaño de efecto de esta comparación es medio, dado que se observa una d de Cohen de 0,305. La diferencia entre el particular pagado y el particular subvencionado fue de 6,977 (IC 95% [3,847, 10,107]); en este caso, el tamaño de efecto también es medio dado que la d de Cohen es de 0,504. Por último, la diferencia media entre el particular pagado y el municipal fue de 11,334 (IC 95% [8,318, 14,351]); al igual que en la anterior, el tamaño de efecto de esta comparación es medio al tomar en cuenta que la d de Cohen es de 0,765.

Por último, se observó homogeneidad de las varianzas en el indicador de IVPT. El análisis ANOVA permitió identificar diferencias significativas entre las tres dependencias ($F(2, 748) = 26,46, p < 0,001$) con un tamaño de efecto mediano ($\eta^2 = 0,066$). Al realizar el análisis post hoc usando la prueba de Tukey, se observa que todas las comparaciones son significativas el nivel de $p < 0,001$, con excepción de la comparación entre el nivel municipal y el particular subvencionado, que es significativa al nivel de 0,026. La diferencia media entre el nivel particular subvencionado y municipal fue de 3,277 (IC 95% [0,307, 6,248]). El tamaño de efecto de esta comparación es pequeño, dado que se observa una d de Cohen de 0,227. La diferencia entre el particular pagado y el particular subvencionado fue de 6,148 (IC 95% [2,977, 9,320]); en este caso, el tamaño de efecto es medio dado que la d de Cohen es de 0,420. Por último, la diferencia media entre el particular pagado y el municipal fue de 9,425 (IC 95% [6,369, 12,482]); al igual que en la anterior, el tamaño de efecto de esta comparación es medio al tomar en cuenta que la d de Cohen es de 0,651.

En la Tabla 13 se muestran las diferencias por género en los indicadores de WISC-V. No se observaron diferencias significativas entre hombres y mujeres en CIT ($F(1, 721) = 0,00, p = 0,960$), ICV ($F(1, 725) = 0,12, p = 0,737$), IRF ($F(1, 746) = 0,00, p = 0,996$), IVE ($F(1, 746) = 0,19, p = 0,662$) e IMT ($F(1, 746) = 0,58, p = 0,449$). Solo en IVP se observó una diferencia significativa entre hombres ($F(1, 746) = 14,98, p < 0,001$), aunque el tamaño de efecto es pequeño ($\eta^2 = 0,020$). La diferencia entre hombres y mujeres en este índice fue de 4,221 (IC 95% [2,082, 6,359]).

Tabla 13
Comparación de Resultados en WISC-V por Sexo

Género		CIT	ICV	IRF	IVE	IMT	IVP
Mujeres	Media	100,0	99,8	100,0	99,8	99,2	102,1
	DE	15,2	15,2	15,0	15,0	15,1	15,1
Hombres	Media	100,0	100,2	100,1	100,3	100,1	97,9
	DE	14,8	14,9	15,0	15,1	15,0	14,6

Limitaciones del Estudio

Es importante considerar algunas limitaciones del presente estudio. En primer lugar, el uso de la dependencia escolar como proxy del NSE es algo que merece atención especial. Si bien este proxy se ha utilizado en estandarizaciones previas (Ramírez & Rosas, 2007; Rosas et al., 2014), es importante notar que la relación no es completamente equivalente. Una segunda limitación guarda relación con el modelo factorial confirmatorio. El objetivo principal de esta investigación fue poner a prueba el modelo factorial de la versión norteamericana para poder contar con una equivalencia internacional del instrumento. Sin embargo, el modelo original ha recibido variadas críticas. Fenollar-Cortés y Watkins (2019) y Watkins et al. (2018) analizaron el modelo con los datos de estandarización de la versión española y canadiense, respectivamente, y concluyen que solo el CIT debería utilizarse para la interpretación de resultados, mientras que la comparación entre índices debería realizarse con extremo cuidado. Ambos grupos de investigadores sugieren modelos alternativos para la estructura de la batería. Una tercera crítica que se ha hecho al modelo factorial de WISC-V tiene relación con el número de factores. Los análisis de Dombrowski et al. (2018) y Canivez et al. (2019) con los datos de la estandarización estadounidense e inglesa, respectivamente, sugieren que el modelo de cuatro factores tiene en general mayor potencia explicativa para la mayoría de los grupos etarios, con algunas excepciones. Estas críticas hacen necesaria una revisión sobre el funcionamiento del modelo factorial en la muestra chilena, la que será presentada en futuras publicaciones. Una tercera limitación del

estudio tiene relación con la decisión de no adaptar los ítems no verbales. Si bien algunas investigaciones han mostrado que este tipo de pruebas tiene una menor sensibilidad cultural (Elkin-Frankston et al., 2007) y que los ítems no verbales son los mismos en todas las versiones internacionales de esta batería, hubiese resultado conveniente hacer un pilotaje completo sobre el funcionamiento de estos ítems. Pese a lo anterior, los resultados psicométricos de estas pruebas dan cuenta de excelentes grados de confiabilidad. Todas estas limitaciones deben tenerse en cuenta para las siguientes estandarizaciones de instrumentos en el país.

Conclusiones y Discusión

Los resultados de la estandarización de WISC-V permiten afirmar que este instrumento cuenta con evidencia de confiabilidad y validez que habilitan su uso en contextos clínicos y educacionales chilenos. En esta línea, y siguiendo las recomendaciones de la American Educational Research Association, American Psychological Association y National Council on Measurement in Education (2014) e Instituto Tecnológico de Chile (Bartram et al., 2018), se vuelve necesario utilizar esta nueva herramienta en reemplazo de WISC-IIIv.ch, dadas las mejoras técnicas y la incorporación de los últimos avances en teorías de evaluación, desarrollo y aprendizaje que incluye WISC-V. El mayor número de indicadores, índices y comparaciones posibles permite obtener más información sobre el funcionamiento cognitivo del evaluado. Estos insumos son particularmente relevantes para realizar apoyos personalizados que tengan el propósito de asistir las necesidades de una intervención educativa remedial.

Respecto a las diferencias por NSE, los resultados de WISC-V muestran variaciones interesantes respecto de WISC-IIIv.ch. En esta última batería, las medias de puntaje por NSE son: municipal = 90,12, particular subvencionado = 99,40 y particular pagado = 110,49 (Rosas & Santa Cruz, 2013). En consecuencia, en WISC-IIIv.ch la diferencia de puntajes en CIT entre el nivel particular pagado y el municipal era de 20,37 puntos estándar, esto es, alrededor de 1,3 desviaciones estándar. En WISC-V, la diferencia entre estas dependencias en el indicador general es de 15,23 (alrededor de 1 desviación estándar), lo que da cuenta de una reducción de 5,14 puntos entre estos niveles. Pese a lo anterior, es importante notar que las brechas entre los niveles siguen siendo significativas y que las diferencias entre el nivel particular pagado y el municipal tienden a ser de entre 0,6 a 1 *DE*. El índice que muestra las menores diferencias por NSE es IVP, que es, posiblemente, el que se ve menos afectado por la calidad de la educación recibida.

Por último, la comparación de resultados por sexo también da cuenta de una reducción en la diferencia entre puntajes de hombres y mujeres. En WISC-IIIv.ch se identificaron diferencias en las tres escalas (CIT, CIE y CIV), aunque con un tamaño de efecto menor (Garolera & Navarro, 2020). En contraste, en WISC-V no se observan diferencias por género, a excepción de la observada en el IVP en favor de las mujeres. Estos resultados son consistentes con otras investigaciones en las que se han observado mejores desempeños en mujeres que en hombres en tareas asociadas a la velocidad de procesamiento (Burns & Nettelbeck, 2005; Camarata & Woodcock, 2006).

En síntesis, la presente investigación muestra que WISC-V puede ser utilizado confiablemente para procesos de evaluación cognitiva en contextos clínicos y educacionales en Chile. La red de investigación de muchas universidades colaboradoras que se ha desarrollado para estos proyectos nacionales de estandarización ha permitido, además, asegurar una rápida diseminación del instrumento a lo largo del país. Se espera contar con más instituciones socias para futuros proyectos de adaptación y estandarización de instrumentos, ya que estos proyectos colaborativos redundan en una mejora clara de los procesos de investigación y formación profesional en las universidades participantes.

Referencias

- Adriasola, A., Ascencio, S., Bernaldes, M., Carrasco, C., Keller, A., Polanco, F., Salas, C. & Santelices, R. (1976). *Estandarización de la Escala Revisada de Wechsler para la medición de la inteligencia en los niños (WISC-R)* [Memoria de grado no publicada]. Pontificia Universidad Católica de Chile.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *The standards for educational and psychological testing*.
- Bartram, D., Berberoglu, G., Grégoire, J., Hambleton, R., Muniz, J. & van de Vijver, F. (2018). ITC guidelines for translating and adapting tests (second edition). *International Journal of Testing*, 18(2), 101-134. <https://doi.org/10.1080/15305058.2017.1398166>
- Burns, N. R. & Nettelbeck, T. (2005). Inspection time and speed of processing: Sex differences on perceptual speed but not IT. *Personality and Individual Differences*, 39(2), 439-446. <https://doi.org/10.1016/j.paid.2005.01.022>

- Calderón, M., Castillo, M., Mandujano, L., Pérez, C. & Purcell, C. (1980). *Estandarización de la Escala Revisada de Inteligencia de Wechsler para niños chilenos del Área Metropolitana (WISC-R). Tercera etapa* [Memoria de grado no publicada]. Pontificia Universidad Católica de Chile.
- Camarata, S. & Woodcock, R. (2006). Sex differences in processing speed: Developmental effects in males and females. *Intelligence*, 34(3), 231-252. <https://doi.org/10.1016/j.intell.2005.12.001>
- Campazzo, E. (1962). *Estandarización de la escala de Wechsler para la medición de la inteligencia en los niños (WISC)* [Memoria de grado no publicada]. Pontificia Universidad Católica de Chile.
- Canivez, G. L., Watkins, M. W. & McGill, R. J. (2019). Construct validity of the Wechsler Intelligence Scale For Children — Fifth UK edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*, 89(2), 195-224. <https://doi.org/10.1111/bjep.12230>
- Cañas, M., Colzani, M., Domínguez, A., Jorquera, K., Orpinas, P. & Valdívieso, P. (1978). *Adaptación, análisis estadístico y construcción de normas de la Escala Revisada de Inteligencia de Wechsler para niños entre 6 años 0 días y 9 años 11 meses 30 días (WISC-R)* [Memoria de grado no publicada]. Pontificia Universidad Católica de Chile.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Crocker, L. & Algina, J. (1986). *Introduction to classical & modern test theory*. Holt, Rinehart & Winston.
- Davies, M. (Comp.) (2008). *The corpus of contemporary American English (COCA)*. [Conjunto de datos]. Brigham Young University. <https://www.english-corpora.org/coca/>
- Dombrowski, S. C., Canivez, G. L. & Watkins, M. W. (2018). Factor structure of the 10 WISC-V primary subtests across four standardization age groups. *Contemporary School Psychology*, 22, 90-104. <https://doi.org/10.1007/s40688-017-0125-2>
- Elkin-Frankston, S., Lebowitz, B. K., Kapust, L. R., Hollis, A. M. & O'Connor, M. G. (2007). The use of the Color Trails Test in the assessment of driver competence: Preliminary report of a culture-fair instrument. *Archives of Clinical Neuropsychology*, 22(5), 631-635. <https://doi.org/10.1016/j.acn.2007.04.004>
- Fenollar-Cortés, J. & Watkins, M. W. (2019). Construct validity of the Spanish version of the Wechsler Intelligence Scale for Children fifth edition (WISC-V^{Spain}). *International Journal of School & Educational Psychology*, 7(3), 150-164. <https://doi.org/10.1080/21683603.2017.1414006>
- Fija normas para determinar los alumnos con necesidades educativas especiales que serán beneficiarios de las subvenciones para educación especial, Decreto 170, Ministerio de Educación de Chile. (2009). https://especial.mineduc.cl/wp-content/uploads/sites/31/2018/06/DTO-170_21-ABR-2010.pdf
- Fuica, P., Lira, J., Alvarado, K., Araneda, C., Lillo, G., Miranda, R., Tenorio, M. & Pérez-Salas, C. P. (2014). Habilidades cognitivas, contexto rural y urbano: comparación de perfiles WAIS-IV en jóvenes. *Terapia Psicológica*, 32(2), 143-152. <https://doi.org/10.4067/S0718-48082014000200007>
- Garolera, M. & Navarro, V. (2020). *Género y WISC en Chile: ¿mejora en el rendimiento o en la evaluación?* (Papeles de Investigación N° 13). Pontificia Universidad Católica de Chile, Escuela de Psicología, Centro de Desarrollo de Tecnologías de Inclusión. <http://descargas.cedeti.cl/2020/06/GENERO%20Y%20WISC.pdf>
- Guilford, J. P. (1954). *Psychometric methods* (2ª ed.). McGraw-Hill.
- Haertel, E. H. (2006). Reliability. En R. L. Brennan (Ed.), *Educational measurement: Fourth edition* (pp. 65-110). Praeger.
- Li, H., Rosenthal, R. & Rubin, D. B. (1996). Reliability of measurement in psychology: From Spearman-Brown to maximal reliability. *Psychological Methods*, 1(1), 98-107. <https://doi.org/10.1037/1082-989X.1.1.98>
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., Bishop, S. L., Luyster, R. J. & Guthrie, W. (2012). *(ADOS™-2) Autism Diagnostic Observation Schedule™, Second edition*. Western Psychological Services.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory: Third edition*. McGraw-Hill.
- Pérez, E., Medrano, L. & Sánchez Rosas, J. (2013). El path analysis: conceptos básicos y ejemplos de aplicación. *Revista Argentina de Ciencias del Comportamiento*, 5(1), 52-66. <https://www.redalyc.org/pdf/3334/333427385008.pdf>
- Ramírez, V. & Rosas, R. (2007). Estandarización del WISC-III en Chile: descripción del test, estructura factorial y consistencia interna de las escalas. *Psykhé*, 16(1), 91-109. <https://doi.org/10.4067/S0718-22282007000100008>
- Rodríguez, M., Rosas, R. & Pizarro, M. (2019). *Rendimiento en escala WISC-V en población urbana y rural de Chile* (Papeles de Investigación N° 11). Pontificia Universidad Católica de Chile, Escuela de Psicología, Centro de Desarrollo de Tecnologías de Inclusión. <http://descargas.cedeti.cl/2019/05/N°11-WISC-URBANO-RURAL.pdf>
- Rosas, R. & Santa Cruz, C. (2013). *Dime en qué colegio estudiaste y te dire qué CI tienes: radiografía al desigual acceso al capital cognitivo en Chile*. Ediciones Universidad Católica de Chile.
- Rosas, R., Tenorio, M., Pizarro, M., Cumsille, P., Bosch, A., Arancibia, S., Carmona-Halty, M., Pérez-Salas, C. P., Pino, E., Vizcarra, B. & Zapata-Sepúlveda, P. (2014). Estandarización de la Escala Wechsler de Inteligencia para Adultos-cuarta edición en Chile. *Psykhé*, 23(1), Artículo 1. <https://doi.org/10.7764/psykhe.23.1.529>
- Sadowsky, S. & Martínez-Gamboa, R. (2012). *Lifcach 2.0: Word frequency list of Chilean Spanish (lista de frecuencias de palabras del castellano de Chile), version 2.0*. Zenodo. <https://doi.org/10.5281/zenodo.268043>
- Silver, N. C. & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology*, 72(1), 146-148. <https://doi.org/10.1037/0021-9010.72.1.146>
- Strube, M. J. (1988). Some comments on the use of magnitude-of-effect estimates. *Journal of Counseling Psychology*, 35(3), 342-345. <https://doi.org/10.1037/0022-0167.35.3.342>
- Sun, J. (2005). Assessing goodness of fit in confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development*, 37(4), 240-256. <https://doi.org/10.1080/07481756.2005.11909764>
- Watkins, M. W., Dombrowski, S. C. & Canivez, G. L. (2018). Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children—Fifth edition. *International Journal of School & Educational Psychology*, 6(4), 252-265. <https://doi.org/10.1080/21683603.2017.1342580>
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children: Fourth edition*. Pearson.
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth edition: Technical and interpretive manual*. Pearson.
- Zhu, J. & Chen, H. -Y. (2011). Utility of inferential norming with smaller sample sizes. *Journal of Psychoeducational Assessment*, 29(6), 570-580. <https://doi.org/10.1177/0734282910396323>

Fecha de recepción: Julio de 2020.
Fecha de aceptación: Abril de 2021.