



Género y WISC en Chile: ¿mejora en el rendimiento o en la evaluación?

Marion Garolera^a y Valentina Navarro^a

^a Centro de Desarrollo de Tecnologías de Inclusión (CEDETi UC). Escuela de Psicología. Pontificia Universidad Católica de Chile.

Resumen

En el campo del estudio del rendimiento cognitivo e inteligencia, persiste el interés por estudiar las diferencias de rendimiento entre los géneros (femenino y masculino), donde los resultados han variados a lo largo del tiempo y en diferentes poblaciones. Los datos normativos de la primera versión del WISC, mostraban un mejor rendimiento en el CIT en niños versus las niñas de forma significativa, pero esta diferencia ha disminuido, e incluso eliminado, en los datos normativos de versiones más actualizadas. En el presente estudio se realiza una comparación por

género a partir de los datos normativos correspondientes a la estandarización de las dos últimas versiones del WISC para población chilena: el WISC-III.v.ch y el WISC-V. Los resultados en WISC-III muestran diferencias significativas en CIT, CIV y CIE, además de 7 subpruebas a favor de los hombres, y solo Claves a favor de las mujeres. En el WISC-V sólo se observan diferencias en IVP a favor de las mujeres, y pequeñas diferencias a nivel de subpruebas en ambos géneros. Las diferencias por género mostradas por el WISC-III.v.ch desaparecen en el WISC-V, lo que indica una disminución del sesgo de construcción del constructo y la prueba. Otras hipótesis para este fenómeno se discuten.

Introducción

La comparación en rendimiento cognitivo por género ha sido un aspecto constantemente abordado en la historia de la evaluación de inteligencia, la cual no ha estado exenta de polémica. Tanto a nivel cognitivo, como a nivel educacional, la diferencia en el rendimiento entre hombres y mujeres continúa siendo un campo de estudio que presenta resultados variados y, muchas veces, contradictorios en una amplia variedad de dominios. Específicamente, el término de comparación en el campo de la inteligencia tiene implicancias que traspasan la evaluación meramente cognitiva al comparar géneros. La definición conceptual de si un sexo es más inteligente que el otro, genera importantes efectos sociopolíticos como la educación compensatoria en el caso de la educación, de medidas de equidad salarial en el ámbito económico (Halpern, 2012).

Ante todo, será relevante considerar que el constructo evaluado definido como “inteligencia” aún no tiene una definición común y generalizable en el campo de la psicología (Bouchard, Boykin,

y Sternberg, 1996; Wasserman, 2012). Debido a esto, el uso de los instrumentos utilizados para su evaluación y el marco teórico en el cual se sostienen estos, serán aspectos centrales que considerar a la hora de observar los resultados. En otras palabras, la limitación sobre la evaluación de inteligencia radica, precisamente, en la limitación de su definición, la cual, a su vez, depende del instrumento que se utilice para evaluar el constructo.

Aparte de la falta de acuerdo en la definición, también existen divergencias en cuanto a la existencia del concepto de *g*, el factor general de inteligencia. Las pruebas que operan en base a la existencia de *g* entregan resultados que permiten resumir el rendimiento intelectual de la persona, como lo es Coeficiente Intelectual Total (CIT). Actualmente existe consenso en que *g* es un indicador sumamente confiable, que se ha mantenido vigente a lo largo del tiempo y resulta ser un determinante importante en diferentes trayectorias de desarrollo social, por lo que es uno de los indicadores más utilizados para referirse al desarrollo intelectual (Lubinski, 2004; Reeve y Charles, 2008; Squalli y Wilson, 2014). Los avances en investigación basados en la relación que tiene *g* con el neurodesarrollo han permitido llegar al consenso de que la inteligencia depende de diferentes procesos cognitivos, y estos dependen de factores genéticos y ambientales, y, por ende, también dependen a su vez de las mediaciones culturales adecuadas (Nisbett et al., 2012a; Sternberg, 2012).

La forma de evaluar inteligencia ha cambiado a medida que avanza la investigación en cognición, neurociencia, biología, desarrollo y evaluación, con el objetivo de lograr resultados que más precisos y socialmente justos. Como los instrumentos han cambiado, al igual que la forma de obtener resultados y su interpretación, es que no se pueden aislar las conclusiones obtenidas en investigaciones anteriores de su periodo histórico y contexto cultural (Sternberg, 2012; Wasserman,

2018). Este punto cobra vital importancia al analizar los estudios que incluyen la comparación de grupos en función de diversos grupos, en el desarrollo cognitivo como puede ser la etnicidad, nivel socioeconómico o el género, especialmente al considerar la importancia que hoy tienen aspectos a nivel sociocultural, como el acceso igualitario a educación y la democratización de estudios universitarios, (Halpern, 2012).

Antecedentes en evaluación de inteligencia y género

La investigación asociada a la comparación de rendimiento intelectual entre hombres y mujeres es de vital importancia para poder identificar si efectivamente, existen diferencias a nivel cognitivo y poder responder a la fuente de éstas diferencias (Halpern, 2012; Nisbett et al., 2012b). Debido a la inexorable relación que existe entre factores biológicos y psicosociales en el desarrollo intelectual y cognitivo, junto con el aprendizaje y el desempeño escolar, es que lograr identificar la razón de esa diferencia a nivel cognitivo puede afectar en la forma en que entendemos cómo cada estudiante responde a las tareas en la evaluación y como aprende a nivel escolar. En ese sentido, entender una diferencia de rendimiento por género implica entender como el contexto y las características de un estudiante pueden afectar su rendimiento en pruebas. Es por esto que, en primer lugar, la diferencia de rendimiento no implica una valoración de un género por sobre otro, y, en segundo lugar, que el desarrollo intelectual no puede observarse de forma aislada sin considerar la complejidad de factores sociales, culturales y académicos.

Dentro de los factores biológicos, existen ciertas diferencias anatómicas a nivel cerebral en cuanto a tamaño, siendo en promedio el cerebro masculino entre un 8% y un 14% más grande que

el del cerebro femenino (Sarikouch et al., 2010), y en cuanto a distribución y concentración de materia gris y blanca en diferentes regiones del cerebro (Eliot, 2011). Sobre este punto, en estudios que asocian los diferentes patrones de concentración de sustancia blanca y gris en hombres y mujeres con resultados en pruebas de inteligencia, muestran resultados similares en CIT (Haier et al 2005). Esto sugiere que diferentes tipos de estructura, diseño y configuración cerebral pueden llegar a un resultado similar en rendimiento intelectual, pero por procesos diferentes. Que mujeres y hombres lleven procesos diferentes ha llevado a líneas de investigación que sugieren que hombre y mujeres difieren en la forma de pensar o de aprender. Sin embargo, otros hallazgos muestran que estas diferencias en el procesamiento a nivel visual, auditivo, en aprendizaje o respuesta al estrés no resultan ser significativas como para sustentar que hombres y mujeres son “pensadores” sustancialmente diferentes (Eliot, 2011).

Por otro lado, estudios que han comparado el rendimiento intelectual y académico de niños y niñas, no han podido hallar diferencias relevantes a nivel global, pero sí se han encontrado a nivel académico en diferentes asignaturas a favor de las mujeres (Deary, Strand, Smith, y Fernandes, 2007). Evaluaciones con pruebas estandarizadas de competencias académicas indican que el género masculino obtiene un rendimiento más alto en las habilidades matemáticas en comparación con las mujeres, quienes obtienen un mejor rendimiento en tareas de expresión oral y escrita (Downing, Chan, Kwong, Lam, y Downing, 2008). Sin embargo, si bien estas diferencias pueden persistir, no ocurre lo mismo con la magnitud de la diferencia en cuanto al rendimiento matemático. Por ejemplo, resultados del examen de aptitud SAT en Estados Unidos durante los 80 mostraban que la distribución de puntajes sobre 700 puntos de hombres (top 0.01%) a mujeres equivalía a la

proporción de 13 a 1 respectivamente (Benbow y Stanley, 1983). Esta brecha ha disminuido significativamente en los últimos años, e incluso ha implicado el aumento de mujeres que obtienen puntajes en el 5% superior en razonamiento verbal y pruebas de expresión escrita (Wai, Cacchio, Putallaz, y Makel, 2010). Este aspecto enfatiza la necesidad de identificar aspectos socioculturales como factores determinantes en los porcentajes de rendimiento de diferentes áreas cognitivas y académicas al comparar el rendimiento de hombres y mujeres en diferentes momentos históricos.

Existen otros factores estudiados asociados a aspectos de personalidad o creencias sobre el rendimiento cognitivo que pueden afectar la variabilidad entre un género y otro, sobre todo considerando ciertos contextos socioculturales que tienden a beneficiar las oportunidades de estimulación a uno de ambos géneros. Estudios han mostrado que se observan diferencias en la autopercepción de rendimiento entre hombres y mujeres, donde los hombres tienden a sobreestimar su rendimiento en el CIT de forma significativa, mientras que las mujeres lo tienden a subestimar levemente (Downing et al., 2008; Reilly y Mulhern, 1995). Sin embargo, al comparar el promedio de CIT en ambos grupos, no se observan diferencias significativas, ni tampoco se aprecia un efecto en cuanto a las creencias de rendimiento en los resultados de las pruebas.

WISC y género

Las escalas Wechsler de inteligencia son una de las pruebas más ampliamente utilizadas alrededor del mundo, aumentando su popularidad gracias a la alta evidencia de óptimas características psicométricas y su relevancia práctica en la aplicación clínica (Groth-Marnat, 2009). Por esta razón, la comparación de rendimiento por género a partir de la escala es uno de los temas

más ampliamente evaluados desde su primera versión, la batería WISC desarrollada el año 1949, hasta su versión actual, el WISC-V. Los resultados en los estudios de estandarización de las diferentes versiones desarrolladas en distintos países distan entre sí. Sin embargo, podemos encontrarnos con ciertas tendencias generales que muestran diferencias significativas, pero pequeñas o estadísticamente nulas a nivel global en el CIT y diferencias más marcadas a nivel de ciertas subpruebas (Kaufman y Doppelt, 1976; Lynn y Mulhern, 1991; Quereshi y Seitz, 1994; Seashore, Wesman, y Doppelt, 1950; Smith, Edmonds, y Smith, 1989) cuyas brechas disminuyen en las últimas dos versiones (H. Chen, Zhang, Raiford, Zhu, y Weiss, 2015; H. Chen y Zhu, 2008; Goldbeck, Daseking, Hellwig-Brida, Waldmann, y Petermann, 2010). Estos resultados son similares a lo observado en otras pruebas que evalúan funcionamiento cognitivo general (Ardila, Rosselli, Matute, y Inozemtseva, 2011; Johnson y Bouchard, 2007).

Dentro de los análisis realizados con los datos de la estandarización de la prueba original, el WISC (Wechsler, 1949), se observan diferencias significativas en los resultados obtenidos por los niños y niñas de la muestra. Específicamente se observa que los hombres presentan un rendimiento más alto en la Escala Verbal (CIV) y en la Escala Total (CIT) con respecto a las mujeres, el cual se presenta de forma estadísticamente significativa (Seashore et al., 1950). Sin embargo, los autores mencionan que estas diferencias resultan ser mínimas, y para fines prácticos resultan ser insignificantes. Frente a estas diferencias, los autores sugieren tres posibles hipótesis. La primera es que realmente los niños posean un mejor rendimiento que las niñas en cuanto a desempeño cognitivo, debido a esto se presentan estas diferencias. Una segunda hipótesis que tiene relación con que el tipo de ítem que se utiliza en las subpruebas tiende a beneficiar a los niños por sobre las niñas.

Finalmente, una tercera hipótesis tiene relación con posibles errores en el muestreo, asociado a que se seleccionaron a niños que pudiesen presentar un rendimiento más alto en comparación con las niñas que se incluyeron en la estandarización. Con los datos disponibles no era posible responder a ninguna de estas hipótesis, pero al menos se abrió el campo para indagar sobre posibles sesgos que pudiesen afectar el rendimiento en pruebas cognitivas.

Con los datos de la estandarización en Estados Unidos del WISC-R (Wechsler, 1974), se mantuvo un mejor rendimiento en hombres a nivel de CIV y CIT, pero la diferencia no sería significativa (Kaufman y Doppelt, 1976). De hecho, lo que antes se presentaba con un mayor nivel de significancia, al analizar los resultados por sexo y por edad, ya no se replicaban en la versión revisada de la batería. Un hallazgo que es importante destacar, es que, aunque las diferencias a nivel de factores eran pequeñas, sí se observaban diferencias significativas de mayor magnitud a nivel de subpruebas. En este aspecto, los resultados de la estandarización muestran que los hombres obtuvieron un rendimiento más alto en la mayoría de las subpruebas, con excepción de la subprueba Claves. En esta última, es dónde se encuentra una mayor distancia entre ambos géneros a favor de las mujeres, quienes alcanzan media desviación estándar de diferencia (Jensen y Reynolds, 1983). A favor de los hombres, se observa una diferencia de un 0.32 de desviación estándar, en la subprueba Información.

Otros estudios con WISC-R en diferentes países muestran resultados similares. Por ejemplo, en los datos de la estandarización para la población galesa e inglesa, se obtiene una diferencia significativa y pequeña a favor de los hombres en CIV y en CIT, no así en CIE (Lynn y Mulhern, 1991). Mayores diferencias se encuentran a nivel de subpruebas, donde los hombres puntúan más

alto en Información, Analogías, Vocabulario, Comprensión, Completación de Figuras, Construcción con Cubos y Ensamblaje de Objetos. En el caso de las mujeres, se observa un mejor rendimiento en Claves y Retención de Dígitos. En general, se repite la tendencia de un mejor rendimiento en los niños en comparación con las niñas en subpruebas y escalas, incluyendo el CIE en algunos estudios, siendo la única excepción la subprueba Claves (Smith et al., 1989).

En el WISC-III los resultados mantienen una tendencia a presentar diferencias leves en los factores más grandes, mientras que la mayor diferencia se encuentra a nivel de subpruebas o dominios específicos (Glutting, Oakland, y Konold, 1994; Weiss y Prifitera, 1995). Los resultados varían entre las diferentes poblaciones, por ejemplo, en los datos de la estandarización de Taiwán se observa un mejor rendimiento en los hombres en tareas de organización perceptual y aritmética, mientras que en las mujeres un mejor rendimiento en retención de dígitos y tareas de velocidad de procesamiento. También se observa una tendencia a que las diferencias por género tienden a disminuir a medida que aumenta la edad (H. Y. Chen, Lynn, y Cheng, 2016). Aunque la mayoría de los estudios muestran diferencias que se presentan como significativas, pero con un bajo tamaño del efecto, igual persisten interpretaciones que atribuyen una diferencia importante en los indicadores generales, sugiriendo incluso, que deberían utilizarse normas diferentes en niños y niñas (Slate, 1998).

Estudios fuera del WISC, muestran resultados similares en las otras pruebas pertenecientes a la familia de las escalas Wechsler de inteligencia, como en la versión para población preescolar y para adultos (WPPSI y WAIS, respectivamente). En el caso de la prueba WPPSI, se observan CITs más altos en el CIT, CIV y CIE en el caso de los niños, existiendo diferencias significativas pero

pequeñas. En esta prueba se aprecian mayores diferencias a nivel de subpruebas, donde Vocabulario (9,42 y 8,03) y Comprensión (10,00 y 8,56) presentan mayores diferencias, a favor de los hombres (Quereshi y Seitz, 1994). En WAIS también se han encontrado hallazgos similares, caracterizado por diferencias menores a nivel general y diferencias más marcadas a nivel de subpruebas (Colom, García, Juan-Espinosa, y Abad, 2002).

Independiente de que los resultados muestran diferencias mínimas en los indicadores globales, surgieron estudios que buscan analizar un posible sesgo en el contexto de evaluación, especialmente al evaluar determinados grupos sociales. Por ejemplo, evaluar la posibilidad de que el evaluador podría generar un efecto en el rendimiento del evaluado dependiendo de su género, nivel socioeconómico o etnia en cuanto a su percepción de motivación y cooperación en la evaluación. Los resultados muestran que no existen diferencias significativas por género, siendo otras variables, como el nivel socioeconómico aquellas que generaban un mayor efecto del sesgo (Glutting et al., 1994; Oakland y Glutting, 1990).

Los datos de estandarización de las últimas versiones de la batería como el WISC-IV y el WISC-V, muestran una nula diferencia entre hombres y mujeres, pero persisten diferencias a nivel de subprueba (H. Chen, Chen, Chang, Lee, y Chen, 2010; H. Chen y Zhu, 2008; Konold y Canivez, 2010). En mayor detalle, en los análisis por género dentro de la muestra normativa de la versión original del WISC-V estadounidense muestra en promedio que las niñas puntúan más alto en los índices principales de Memoria de Trabajo y Velocidad de Procesamiento, los índices secundarios No Verbal y de Competencias Cognitivas, además del CIT. Sin embargo, el tamaño de las diferencias era mínimo, con excepción de los índices que involucran velocidad de procesamiento.

Por su parte, los niños obtuvieron un rendimiento significativamente más alto en el Índice de Razonamiento Cuantitativo, mientras que en el Índice Visoespacial se obtuvieron resultados levemente superiores a los obtenidos por las niñas.

En Chile, la única versión estandarizada anterior al WISC-V, el WISC-III no presenta estudios que realicen la comparación por género, por lo que en el presente estudio se realizará una comparación entre esos datos y los recientemente obtenidos en el WISC-V. Se hipotetiza que es probable encontrar diferencias sutiles o nulas en el CIT entre hombres y mujeres de la muestra chilena en ambas versiones y diferencias más marcadas a nivel de subpruebas. También se hipotetiza que de haber diferencias significativas a nivel de CIT entre hombres y mujeres en el WISC-III, estas no se replicarán en la versión del WISC-V.

Metodología

Participantes

La muestra utilizada comprende los datos normativos de las estandarizaciones chilenas para las pruebas WISC-III y WISC-V.

El WISC-III incluye una muestra de 1924 casos, divididos equivalentemente en 963=niñas y 961=niños. Los casos fueron evaluados en 4 regiones de Chile (Coquimbo, Valparaíso, Araucanía, y Metropolitana). Además, hay equivalencia entre dependencias educativas (644= Municipal, 640 Subvencionada, y 640 Particular pagada).

Para el caso de WISC-V la muestra es menor, con 751 casos (387=niñas y 364=niños). Incluye casos de 7 regiones de Chile (Arica y Parinacota, Antofagasta, Ñuble, Bío-Bío, Araucanía

y Los Lagos), y de las tres dependencias escolares (municipal= 286, particular subvencionado=245 y privado=219).

Ambas muestras nacionales fueron divididas en 22 grupos etarios, desde los 6 a los 16 años. Una descripción detallada de ambas muestras normativas puede encontrarse en sus respectivos manuales (Ramírez y Rosas, 2007, Rosas y Pizarro, 2018).

Procedimiento

Los niños fueron seleccionados aleatoriamente de diversos colegios de las regiones incluidas en cada muestra. Sólo los casos en los que se pudieron obtener de manera independiente consentimiento de los padres y asentimiento de los participantes fueron evaluados. Las evaluaciones se realizaron en los establecimientos, o en las dependencias del Centro de Desarrollo de Tecnologías de Inclusión (CEDETi UC), por profesionales capacitados para tal objetivo. Tras la evaluación, cada NNA recibió un regalo simbólico.

Instrumentos

WISC-III:

La estructura del WISC-III presenta un indicador de funcionamiento cognitivo general o CIT, el cual se divide a su vez, en una Escala Verbal (CIV) y una Escala de Ejecución (CIE). Además, se pueden obtener cuatro índices adicionales; el Índice de Comprensión Verbal, el Índice de Organización Perceptual, el Índice de Ausencia de Distractibilidad y el Índice de Velocidad de Procesamiento. El WISC-III en total tiene 13 subpruebas, siendo 10 las subpruebas que permiten

obtener las tres escalas (CIT, CIV y CIE) y 13 subpruebas para obtener los 4 índices adicionales. Los 4 índices se componen por diferentes subpruebas de acuerdo con la habilidad que evalúan, y su composición no es equitativa para todos los índices en cuanto al número de subpruebas que lo componen. Con la aplicación de cada subprueba tiene un puntaje bruto, que luego se transforma a puntajes estándar, usando los datos normativos obtenidos del proceso de estandarización de la prueba WISC-III versión de Chile. Con esos puntajes estandarizados, se calculan los índices, las escalas Verbal y de Ejecución, y el Coeficiente Intelectual Total (CIT).

WISC-V:

El WISC-V se compone de 15 subpruebas, que en base a su aplicación se permite la obtención de 10 indicadores además del CIT. Se incluyen cinco índices principales; el Índice de Comprensión Verbal (ICV), el Índice Visoespacial (IVE), el Índice de Razonamiento Fluido (IRF), el Índice de Memoria de Trabajo (IMT) y el Índice de Velocidad de Procesamiento (IVP). Además, se incorporan cinco índices secundarios; el Índice de Memoria de Trabajo Auditiva (IMTA), el Índice de Razonamiento Cuantitativo (IRC), el Índice No Verbal (INV), el Índice de Habilidad General (IHG) y el Índice de Competencia Cognitiva (ICC).

La estructura del WISC-V presenta tres niveles de análisis, independientes entre sí. En el primer nivel se encuentra el CIT el cual puede ser obtenido tras la aplicación de 7 subpruebas y que representa ser un resumen del funcionamiento cognitivo general. Con 10 subpruebas, se pueden obtener los 5 índices principales, cada uno compuesto por dos subpruebas que evalúan un

determinado dominio cognitivo, y con 13 subpruebas se pueden obtener los 5 índices secundarios que permiten aportar con información adicional, dependiendo del rendimiento del evaluado.

La prueba se compone de 15 subpruebas, cada una tiene un puntaje bruto, que luego se transforma a puntajes, usando los datos normativos obtenidos del proceso de estandarización de la prueba WISC-V versión de Chile. Estos puntajes estandarizados, se calcularon los cinco índices principales y los cinco índices secundarios y el Coeficiente Intelectual Total (CIT).

Para una comprensión más profunda de estas baterías, consultar los manuales técnicos de WISC-III y WISC-V versión chilena (Ramírez y Rosas, 2007; Rosas y Pizarro, 2018).

Análisis de datos

Se obtuvieron los datos descriptivos de las dos muestras totales, de WISC-III y WISC-V, y se obtuvieron datos descriptivos en función del género (femenino y masculino) y la dependencia económica del colegio de él o la evaluada, la variable de Dependencia fue usada como proxy del nivel socioeconómico, al existir una fuerte y consistente correlación entre ambas variables en el contexto chileno (Santacruz y Rosas, 2013). Además, se resumieron los resultados encontrados en ambas baterías.

Estos análisis se realizaron utilizando el software IBM SPSS v.23 (IBM, 2019).

Resultados

Los resultados presentados en las tablas a continuación corresponden a un análisis propio de la muestra normativa de WISC-III y WISC-V (Ramírez y Rosas, 2007; Rosas y Pizarro, 2018).

1) WISC-III

Diferencias a nivel de subprueba

Como se puede observar en la siguiente tabla, las subpruebas en WISC-III muestran una diferencia de género en 8 de las 13 subpruebas. 7 de estas diferencias son a favor del género masculino. Y sólo una, Claves, a favor de las mujeres. Al examinar el tamaño del efecto de estas diferencias (columna Est.), se destaca que Información, Aritmética, Claves y Laberintos son de tamaño pequeño, el resto es mínimo y no se recomienda desprender conclusiones acerca de ellas.

Tabla 1

Puntajes subpruebas WISC-III

Sub prueba	Fem.	Masc.	Sig.	Est.
Información	9.4	10.5	0.00*	0.18
Analogías	9.9	10.1	0.18	0.03
Aritmética	9.3	10.1	0.00*	0.13
Vocabulario	9.8	10.3	0.00*	0.08
Comprensión	10.1	9.9	0.17	0.03
Retención de dígitos	9.9	10.0	0.96	0.00
Completación de Figuras	9.9	10.3	0.37	0.02
Ordenamiento de historias	9.8	10.0	0.00*	0.08
Construcción con Cubos	9.7	10.4	0.00*	0.11
Ensamble de objetos	9.6	10.3	0.00*	0.10
Claves	10.4	9.5	0.00*	0.15
Laberintos	9.6	10.5	0.00*	0.14
Búsqueda de símbolos	10.2	10.0	0.28	0.02

Diferencias a nivel de índice

Se puede observar que para las tres escalas (Verbal, Ejecución y Total) existen diferencias significativas por género, las tres favoreciendo a los hombres. Pero el tamaño del efecto de tales diferencias es mínimo (menor o igual a 0,01).

Respecto al nivel socioeconómico, se observa la presencia de diferencias significativas en los tres índices, y con un tamaño del efecto considerable (sobre 0,2).

En tercer lugar, cabe preguntarse si existe un efecto general de interacción entre las dos variables (género y NSE), y como se observa en la tabla 2, no es posible afirmar la existencia de esta interacción al observar los niveles de significación y tamaño del efecto (bajo el título de interacción).

Tabla 2

Efectos de interacción en las medias de los índices de WISC-III.

Índice	Género				NSE		Interacción	
	Masc.	Fem.	Sig.	Est.	P	Efecto	Sig.	Est.
CI Verbal	101.4	98.6	.000	0.01	.000	0.28	.660	0.00
CI Ejecución	100.7	99.3	.010	0.00	.000	0.22	.360	0.00
CI Total	101.2	98.8	.000	0.01	.000	0.31	.460	0.00

2) WISC-V

Diferencias a nivel de subprueba

A nivel de las subpruebas se pueden encontrar algunas diferencias entre los géneros. No existe un patrón claro de mayor desempeño de un género sobre el otro; en algunas pruebas los niños se desempeñan en promedio mejor y en otras las niñas. Sí existen algunas diferencias significativas observadas: en Retención de Dígitos e Información los niños han tenido un mejor desempeño, así como las niñas han obtenido puntajes promedio mayores en Claves, Búsqueda de Símbolos y Compresión. De todas formas, los tamaños del efecto de estas diferencias son mínimos, por lo que toda conclusión desprendida a partir de estas diferencias debe ser tomada con prudencia.

Diferencias a nivel de índice:

Estas diferencias no se ven reflejadas a nivel de índices. Como indica la tabla 4, se observó en el análisis de los puntajes de los 11 índices que no existiría una diferencia significativa que favorezca a niños o niñas. La excepción se encuentra en el índice de velocidad de procesamiento en el que el género femenino supera en 4 puntos estándares al género masculino, una diferencia que de todas formas tiene un efecto mínimo (0,02).

Tabla 3

Medias de subpruebas en WISC-V.

Subprueba	Femenino	Masculino	Sig.	Est.
CC	10.0	10.2	.392	0.00
AN	9.8	9.7	.645	0.00
MR	10.2	10.0	.412	0.00
RD	10.0	10.4	.041*	0.01
CL	10.7	9.8	.000*	0.03
VO	9.9	9.8	.707	0.00
BL	9.9	10.2	.248	0.00
RV	10.1	10.0	.646	0.00
RI	10.1	10.1	.926	0.00
BS	10.3	9.8	.034*	0.01
IN	9.7	10.3	.002*	0.01
SL	10.0	10.1	.536	0.00
CN	9.9	10.2	.188	0.00
CO	10.2	9.4	.006*	0.01
AR	9.9	10.2	.061	0.01

Tabla 4

Medias de índices en WISC-V.

Índices	Femenino	Masculino	Sig.	Est.
CIT	97.9	95.6	.186	0.00
ICV	98.1	96.2	.233	0.00
IRF	100.1	100.3	.661	0.00
IVE	100.2	100.2	.874	0.00
IMT	99.5	100.5	.247	0.00
IVP	102.5	98.3	.000*	0.02
IMTA	99.6	100.7	.163	0.00
IRC	99.3	101.2	.074	0.00
ICC	101.4	99.5	.125	0.00
IHG	97.6	95.9	.356	0.00
INV	99.6	99.1	.763	0.00

3) Interacción con NSE ¿existe mayor brecha en algún NSE?

Como ya se ha mencionado, a nivel de índice, se encontró sólo una diferencia significativa entre géneros, siendo esta de un tamaño despreciable. Se exploró la presencia de otras diferencias demográficas, como el nivel socioeconómico, variable en la que históricamente existen brechas que favorecen a los NNA de niveles socioeconómicos más altos.

En los resultados del WISC-V se ha encontrado que la brecha por nivel socioeconómico sigue persistiendo, observándose diferencias significativas, que favorecen al nivel socioeconómico alto en todos los índices incluido el CI total. Estas diferencias poseen un tamaño del efecto bajo (0,1 o menos).

De esta manera se observa que existe un efecto principal, respecto a si la brecha de género se agudiza en distintos niveles socioeconómicos, se puede observar que no hay antecedentes para afirmar la existencia de un efecto de interacción; como indica la tabla bajo el título de Interacción, no hay efectos significativos de interacción.

Tabla 5

Efectos de interacción género/nivel socioeconómico en WISC-V.

Índice	Género			NSE		Interacción	
	Sig.	Efecto	Favor	Sig.	Efecto	Sig.	Efecto
CIT	.186	0.00	Fem	.000	0.08	.664	0.00
ICV	.233	0.00	Fem	.000	0.09	.886	0.00
IRF	.661	0.00	Masc	.000	0.07	.169	0.01
IVE	.874	0.00	Masc	.000	0.06	.084	0.01
IMT	.247	0.00	Masc	.000	0.07	.116	0.01
IVP	.000	0.02	Fem	.000	0.05	.681	0.00
IMTA	.163	0.00	Masc	.000	0.07	.156	0.01
IRC	.074	0.00	Masc	.000	0.10	.248	0.00
ICC	.125	0.00	Fem	.000	0.09	.224	0.00
IHG	.356	0.00	Fem	.000	0.08	.754	0.00
INV	.763	0.00	Fem	.000	0.10	.170	0.01

Discusión

¿Diferencias género por subprueba?

Al observar los resultados encontrados en WISC-V se puede resumir que: en la subprueba de Comprensión, se evidencian puntajes mayores en las niñas, lo que podría hablar de unas mayores habilidades de adaptación social en función de la edad, según lo que los expertos han definido como respuestas adaptativas. También es posible que ventajas asociadas a la expresión del lenguaje en niñas permita obtener respuestas más complejas y desarrolladas que podrían beneficiar el puntaje en esta prueba. Este resultado debe interpretarse de manera cauta, al tener un tamaño del efecto estadísticamente despreciable.

Para Claves y Búsqueda de símbolos, se observan puntajes que favorecen a las niñas, lo que indicaría mayor capacidad de atención, procesamiento visual, entre otras. Esto va en la línea de lo encontrado en las otras versiones de la prueba (Chen et al., 2010; Goldbeck et al., 2010; Jensen y Reynolds, 1983; Slate y Jones, 1998), en que la mayor diferencia a nivel de subpruebas estaba en Claves, favoreciendo a las mujeres.

En la subprueba de Información se ha encontrado una diferencia significativa que favorece al género masculino. Smarter (2000) indica que en la Subprueba de Información suele favorecer en puntaje a la población masculina. Este hallazgo puede ser especialmente decidor de posibles sesgos en la construcción al ser una prueba que requiere una selección de contenidos culturalmente relevantes, y al observarse que los puntajes en Información de WISC-V Chile favorece a los hombres, es posible que se hayan dejado fuera de la subprueba contenidos relevantes para niñas.

Pero, este posible sesgo finalmente genera una diferencia sin impacto a nivel teórico, al poseer un tamaño del efecto de 0,01.

Los niños, además, presentan mejores puntajes en Retención de Dígitos, lo que indicaría una mayor capacidad de memoria a corto plazo y memoria de trabajo. Dos habilidades muy relacionadas a la inteligencia fluida, independientes del capital cultural (Diamond, 2013), esto podría implicar la existencia de un tamaño del efecto relevante, lo cual no sucede en los resultados encontrados (Efecto de 0,01). Este hallazgo además llama la atención, ya que históricamente esta prueba había favorecido al género femenino (ver, por ejemplo, las versiones galesa e inglesa de WISC-R).

Las diferencias a nivel de subprueba pueden entregar evidencia de sesgos en la construcción de tales tareas, y como se ha observado, la posible presencia de sesgos de construcción por género es muy baja, y no favorece a un género sobre otro de manera consistente.

¿Y en los índices?

Al “subir” de nivel a los índices, la evidencia permite hacer otro tipo de inferencias, que se acercan a las variables latentes que la batería WISC-V está midiendo. En este nivel es posible observar una diferencia significativa entre géneros sólo en el índice de Velocidad de Procesamiento que favorece a las niñas, lo cual concuerda con los hallazgos presentados en los análisis por género de la versión original del WISC-V. Sin embargo, en el caso de la norma chilena, esta diferencia tiene un tamaño del efecto muy bajo (0,02) por lo que las conclusiones deben ser tomadas con precaución. De igual forma, la composición de los índices a nivel de subprueba tuvo una fuerte variación entre la tercera y la quinta versión de la batería al pasar de cuatro factores con composición

variable en cuanto al número de subpruebas, a una estructura de cinco factores con una composición equitativa de dos subpruebas por cada uno de ellos.

Es importante destacar que esta diferencia palidece ante las similitudes encontradas entre niños y niñas, especialmente relevante es el hallazgo de la equivalencia en CIT, que indica la ausencia de sesgos importantes en la prueba en su conjunto. Este hallazgo va en la línea de numerosos estudios que señalan que hombres y mujeres tienen un rendimiento equiparable en funcionamiento cognitivo general a lo largo de la vida.

¿Qué sucede al comparar WISC-III y WISC-V?

Al comparar los resultados encontrados en la muestra normativa de WISC-V con la versión anterior (WISC-III), se observa que en WISC-III hay diferencias significativas por género en los tres índices, que favorecen a los hombres, pero no se observan diferencias en los índices de WISC-V. La comparación entre ambas baterías permite sacar conclusiones respecto de sus características psicométricas, pero, no necesariamente de su población. Para ello sería necesario realizar un estudio comparativo con ambas pruebas aplicadas a los mismos sujetos o un estudio longitudinal. En esta línea, con los antecedentes presentados se puede inferir que WISC-V ha sido construido y aplicado en terreno con un menor sesgo de género, lo que a su vez se ha traducido en puntajes equivalentes entre género femenino y masculino.

Pero además del ámbito de las características psicométricas, vale la pena preguntarse por posibles efectos del cambio estructural del modelo de rendimiento cognitivo entre una prueba y otra. En el WISC-III, los índices apuntaban a una división entre dos grandes factores representados en el

CI Verbal (CIV) y el CI de Ejecución (CIE), los cuales engloban una mayor cantidad de subpruebas que a la base podían representar habilidades que se encuentran fuera de la división verbal-ejecutiva. De hecho, es este mismo punto que lleva a la eliminación total de la estructura bifactorial para pasar de forma directa a factores que representan dominios cognitivos específicos en el WISC-V, lo cual podría permitir un mejor ajuste psicométrico a la evaluación cognitiva evitando el sesgo por género. Este aspecto también cobra relevancia al considerar el cambio estructural que ha presentado el mismo CIT en ambas versiones, al eliminarse subpruebas con una mayor carga académica como Aritmética e Información, y que tendían a beneficiar a los niños por sobre las niñas en versiones anteriores de la prueba.

Aun así, se puede hipotetizar que este cambio entre una versión y otra es signo de un cambio educacional a nivel nacional que ha disminuido las diferencias entre ambos géneros, por ejemplo, al ocurrir una democratización a nivel internacional del ingreso de hombres y mujeres a la universidad (Halpern, 2012), o el reciente aumento del interés en distintas instituciones del sistema educativo nacional por fomentar una educación no sexista que fomente la erradicación de los estereotipos de género en la escuela (Acuña y Montesinos, 2014). Interés que ha sido expresado por el Colegio de Profesores (2019), y por la agencia de Educación 2020.

En relación con el NSE, se encuentran diferencias significativas por esta variable en ambas versiones, persistiendo así la brecha socioeconómica, estos resultados van en la línea de lo que Rosas y Santa-Cruz (2013) han señalado en su trabajo y lo que Oakland y Glutting (1990;1994) han

mostrado, el género constituye un bajo impacto en los índices en comparación a, por ejemplo, el nivel socioeconómico.

Además, cabe destacar, que no se observa una interacción entre género y NSE, lo que significaría que la diferencia por género tendría un tamaño relativamente parecido en distintos niveles socioeconómico, o, dicho de otra forma, la distancia entre hombres y mujeres no se volvería más aguda al subir o bajar de nivel socioeconómico.

Los resultados encontrados a nivel de grupo en el WISC-V utilizando la norma chilena no concuerdan con los hallazgos observados con la norma estadounidense, a excepción de lo observado en velocidad de procesamiento, y especialmente la subprueba Claves con un rendimiento a favor de las niñas. Este aspecto se ha mantenido de forma constante en las diferentes versiones de la prueba, y en los resultados de diferentes países por lo que es posible concluir que las mujeres obtienen un mejor rendimiento en tareas de velocidad de procesamiento de forma consistente.

Más allá, el largo y extenso salto en años y avances a nivel teórico desde la primera versión propiamente estandarizada del WISC-III al WISC-V en Chile, incorpora cambios a nivel de estructura y formato en la forma de evaluar inteligencia que afecta la posibilidad de realizar análisis comparativos equivalentes. Sin embargo, es posible concluir en el caso de haber existido un sesgo en la construcción de la tercera versión de la prueba, que beneficiara a un género por sobre el otro, este ya no es observable en la quinta versión de la prueba.

Referencias

- Acuña, M. y Montesinos, S. (2014). La otra reforma: El “no sexismo” como clave cultural del cambio en el sistema educacional. *Revista Anales, 1* (7). <https://doi.org/10.5354/0717-8883.2015.35890>.
- Ardila, A., Rosselli, M., Matute, E., y Inozemtseva, O. (2011). Gender Differences in Cognitive Development. *Developmental Psychology, 47*(4), 984–990. <https://doi.org/10.1037/a0023819>
- Benbow, C. P., y Stanley, J. C. (1983). Sex differences in Mathematical Reasoning Ability: More Facts. *Science, 222*, 1029–1031. <https://doi.org/10.1126/science.6648516>
- Bouchard, T. J., Boykin, A. W., y Sternberg, R. J. (1996). Intelligence: Knowns and Unknowns Intelligence: Knowns and Unknowns. *American Psychologist 51*(2):77-101. <https://doi.org/10.1037/0003-066X.51.2.77>
- Chen, H. Y., Lynn, R., y Cheng, H. (2016). Sex differences on the WISC-III in Taiwan and the United States. *Mankind Quarterly, 57*(1), 66–71.
- Chen, H., Chen, M. F., Chang, T. S., Lee, Y. S., y Chen, H. P. (2010). Gender reality on multi-domains of school-age children in Taiwan: A developmental approach. *Personality and Individual Differences, 48*(4), 475–480. <https://doi.org/10.1016/j.paid.2009.11.027>
- Chen, H., y Zhu, J. (2008). Factor invariance between genders of the Wechsler Intelligence Scale for Children - Fourth edition. *Personality and Individual Differences, 45*(3), 260–266. <https://doi.org/10.1016/j.paid.2008.04.008>

-
- Chen, H., Zhang, O., Raiford, S. E., Zhu, J., y Weiss, L. G. (2015). Factor invariance between genders on the Wechsler Intelligence Scale for Children-Fifth Edition. *Personality and Individual Differences*, 86, 1–5. <https://doi.org/10.1016/j.paid.2015.05.020>
- Colegio de Profesores. (2018). *Consideraciones para una educación Educación No Sexista*. Extraído de <https://www.colegiodeprofesores.cl/wp-content/uploads/2018/06/CONSIDERACIONES-PARA-UNA-EDUCACION-NO-SEXISTA.pdf>
- Colom, R., García, L. F., Juan-Espinosa, M., y Abad, F. J. (2002). Null sex differences in general intelligence: Evidence from the WAIS-III. *Spanish Journal of Psychology*, 5(1), 29–35. <https://doi.org/10.1017/S1138741600005801>
- Deary, I. J., Strand, S., Smith, P., y Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1), 13–21. <https://doi.org/10.1016/j.intell.2006.02.001>
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64 (1), 135-168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Downing, K., Chan, S., Kwong, T., Lam, T., y Downing, W. (2008). Measuring gender differences in cognitive functioning. *Multicultural Education y Technology Journal*, 2(1), 4–18. <https://doi.org/10.1108/17504970810867124>
- Eliot, L. (2013). Single-sex education and the brain. *Sex Roles*, 69(7-8), 363-381. <https://doi.org/10.1007/s11199-011-0037-y>

-
- Glutting, J. J., Oakland, T., y Konold, T. R. (1994). Criterion-related bias with the guide to the assessment of test-session behavior for the WISC-III and WIAT: Possible race/ethnicity, gender, and SES effects. *Journal of School Psychology, 32*(4), 355–369.
[https://doi.org/10.1016/0022-4405\(94\)90033-7](https://doi.org/10.1016/0022-4405(94)90033-7)
- Goldbeck, L., Daseking, M., Hellwig-Brida, S., Waldmann, H. C., y Petermann, F. (2010). Sex differences on the German Wechsler Intelligence Test for Children (WISC-IV). *Journal of Individual Differences, 31*(1), 22–28. <https://doi.org/10.1027/1614-0001/a000003>
- Groth-Marnat, G. (2009). *Handbook of psychological assessment*. John Wiley y Sons.
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., y Alkire, M. T. (2005). The neuroanatomy of general intelligence: sex matters. *NeuroImage, 25*(1), 320-327.
<https://doi.org/10.1016/j.neuroimage.2004.11.019>
- Halpern, D. F. (2012). *Sex Differences in Cognitive Abilities* (Fourth). New York: Psychology Press. <https://doi.org/10.4324/9781410605290>
- IBM. (2019). *Downloading IBM SPSS Statistics 23*. Extraído de
<https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-23>
- Jensen, A. R., y Reynolds, C. R. (1983). Sex differences on the WISC-R. *Personality and Individual Differences, 4*(2), 223–226. [https://doi.org/10.1016/0191-8869\(83\)90029-6](https://doi.org/10.1016/0191-8869(83)90029-6)
- Johnson, W., y Bouchard, T. J. (2007). Sex differences in mental abilities: g masks the dimensions on which they lie. *Intelligence, 35*(1), 23–39. <https://doi.org/10.1016/j.intell.2006.03.012>

Kaufman, A. S., y Doppelt, J. E. (1976). Analysis of WISC-R Standardization Data in Terms of the Stratification Variables. *Child Development*, 47(1), 165–171.

<https://doi.org/10.2307/1128295>

Konold, T. R., y Canivez, G. L. (2010). Differential relationships between WISC-IV and WIAT-II scales: An evaluation of potentially moderating child demographics. *Educational and Psychological Measurement*, 70(4), 613–627. <https://doi.org/10.1177/0013164409355686>

Lubinski, D. (2004). Introduction to the Special Section on Cognitive Abilities: 100 Years After Spearman's (1904) "General Intelligence, Objectively Determined and Measured." *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/0022-3514.86.1.96>

Lynn, R., y Mulhern, G. (1991). A comparison of sex differences on the Scottish and American standardisation samples of the WISC-R. *Personality and Individual Differences*, 12(11), 1179–1182. [https://doi.org/10.1016/0191-8869\(91\)90082-M](https://doi.org/10.1016/0191-8869(91)90082-M)

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., y Turkheimer, E. (2012a). Intelligence: New findings and theoretical developments. *American Psychologist*, 67(2), 130–159. <https://doi.org/10.1037/a0026699>

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., y Turkheimer, E. (2012b). Intelligence: New Findings and Theoretical Developments. *American Psychologist*, 67(2), 130–159. <https://doi.org/10.1037/a0026699>

Oakland, T., y Glutting, J. J. (1990). Examiner Observations of Children's WISC-R Test-Related Behaviors: Possible Socioeconomic Status, Race, and Gender Effects. *Psychological*

Assessment, 2(1), 86–90. <https://doi.org/10.1037/1040-3590.2.1.86>

Quereshi, M. Y., y Seitz, R. (1994). Gender differences on the WPPSI, the WISC-R, and the WPPSI-R. *Current Psychology*, 13(2), 117–123. <https://doi.org/10.1007/BF02686795>

Ramírez, V., y Rosas, R. (2007). Estandarización del WISC-III en Chile: Descripción del test, estructura factorial y consistencia interna de las escalas. *Psyche*, 16(1), 91–109. <https://doi.org/10.4067/s0718-22282007000100008>

Reeve, C. L., y Charles, J. E. (2008). Survey of opinions on the primacy of g and social consequences of ability testing: A comparison of expert and non-expert views. *Intelligence*. <https://doi.org/10.1016/j.intell.2008.03.007>

Reilly, J., y Mulhern, G. (1995). Gender differences in self-estimated IQ: The need for care in interpreting group data. *Personality and Individual Differences*, 18(2), 189–192. [https://doi.org/10.1016/0191-8869\(94\)00148-L](https://doi.org/10.1016/0191-8869(94)00148-L)

Rosas, R. y Pizarro, M. (2018). *WISC-V. Manual de Administración y Corrección*. Santiago Chile. CEDETi-UC.

Rosas, R., y Santa Cruz, C. (2013). *Dime en qué colegio estudiaste y te diré qué CI tienes: Radiografía al desigual acceso al capital cognitivo en Chile*. Ediciones UC.

Sarikouch, S., Peters, B., Gutberlet, M., Leismann, B., Kelter-Kloeping, A., Koerperich, H., y Beerbaum, P. (2010). Sex-specific pediatric percentiles for ventricular size and mass as reference values for cardiac MRI: assessment by steady-state free-precession and phase-

contrast MRI flow. *Circulation: Cardiovascular Imaging*, 3(1), 65-76.

<https://doi.org/10.1161/CIRCIMAGING.109.859074>

Seashore, H., Wesman, A., y Doppelt, J. E. (1950). The Standardization of the Wechsler Intelligence Scale for Children. *Journal of Consulting Psychology*, 14, 99–110.

<https://doi.org/10.1037/h0056307>

Slate, J. R. (1998). Sex differences in WISC-III IQs: Time for separate norms? *Journal of Psychology: Interdisciplinary and Applied*, 132(6), 677–679.

<https://doi.org/10.1080/00223989809599300>

Slate, J. R., y Jones, C. H. (1998). Gender Differences and Factor Structure of Wisc-III Scores for Students with Mild Mental Retardation. *Diagnostique*, 23(2), 61–75.

<https://doi.org/10.1177/073724779702300201>

Smith, T. C., Edmonds, J. E., y Smith, B. (1989). The role of sex differences in the referral process as measured by the Peabody Picture Vocabulary Test-Revised and the Wechsler Intelligence Scale for Children-Revised. *Psychology in the Schools*, 26(October), 354–358.

[https://doi.org/10.1002/1520-6807\(198910\)26:4<354::AID-PITS2310260405>3.0.CO;2-A](https://doi.org/10.1002/1520-6807(198910)26:4<354::AID-PITS2310260405>3.0.CO;2-A)

Squalli, J., y Wilson, K. (2014). Intelligence, creativity, and innovation. *Intelligence*, 46, 250–257.

<https://doi.org/10.1016/j.intell.2014.07.005>

Sternberg, R. J. (2012). Intelligence. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(5), 501–511. <https://doi.org/10.1002/wcs.1193>

- Wai, J., Cacchio, M., Putallaz, M., y Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30year examination. *Intelligence*, 38(4), 412–423.
<https://doi.org/10.1016/j.intell.2010.04.006>
- Wasserman, J. D. (2012). A History of Intelligence Assessment: The unfinished tapestry. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues*.
- Wasserman, J. D. (2018). A history of intelligence assessment: The unfinished tapestry. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, 3–55.
- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York, NY: Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4rd ed.). San Antonio, TX: Pearson.
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5rd ed.). Bloomington, MN: Pearson.
- Weiss, L. G., y Prifitera, A. (1995). An evaluation of differential prediction of WIAT achievement scores from WISC-III FSIQ across ethnic and gender groups. *Journal of School Psychology*, 33(4), 297–304. [https://doi.org/10.1016/0022-4405\(95\)00016-F](https://doi.org/10.1016/0022-4405(95)00016-F)