



La importancia de la invarianza métrica para la investigación en sociología

[The importance of measurement invariance for research in sociology]

Daniel Miranda ¹ y Julio Iturra ²

¹Pontificia Universidad Católica de Chile; ²Centro de Estudios de Conflicto y Cohesión Social

Resumen

Comparar variables latentes presenta importantes desafíos en cuanto a su equivalencia. Usando los datos del estudio ICCS 2016, en la que participaron 24 países, evaluamos la adecuación métrica de concepto de actitudes hacia la igualdad de derechos contrastando un modelo unidimensional (disponible en la documentación original) con un modelo bidimensional. Además, evaluamos los niveles de comparabilidad de las mediciones con análisis factorial confirmatorio multigrupo. Los resultados muestran que el modelo bidimensional logra sustanciales mejoras. Adicionalmente, la evaluación de la invarianza del modelo bidimensional muestra adecuados niveles de comparabilidad (invarianza métrica y escalar). Esto quiere decir que el instrumento especificado en sus dos dimensiones posee una comprensión similar por los participantes de los distintos países de la muestra. Por tanto, se podría comparar los niveles de cada dimensión y la relación de estas dimensiones con otras variables.

Palabras clave: invarianza, análisis factorial confirmatorio, actitudes hacia la igualdad de género

Abstract

The comparison of latent variables presents considerable challenges in terms of their equivalence. Using data from ICCS 2016, in which 24 countries participated, we assess the measurement adequacy of the concept of attitudes towards equal rights by contrasting a one-dimensional model (available in the original documentation) with a two-dimensional model. In addition, we evaluate the levels of comparability of the measurements using multigroup confirmatory factor analysis. Results show that the two-dimensional model achieves substantial improvements. Additionally, evaluation of the invariance of the two-dimensional model shows adequate levels of comparability (metric and scalar invariance). This means that the instrument specified in its two dimensions permits a similar understanding by the participants in the various countries included in the sample. As such, it would be possible to compare both the levels of each dimension and the relationship of these dimensions with other variables.

Keywords: invariance, confirmatory factor analysis, attitudes towards equal rights

Contacto: La comunicación sobre este artículo debe ser enviada a Daniel Miranda, email damiran1@uc.cl

Financiamiento: Esta investigación contó con financiamiento de la Agencia Nacional de Investigación y Desarrollo (ANID) a través del Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT 11190508) y el Fondo de Financiamiento de Centros de Investigación en Áreas Prioritarias (FONDAP 15130009).



INTRODUCCIÓN

Comparar grupos, contextos o sociedades es un constitutivo de la disciplina Sociológica. Desde una aproximación cuantitativa, para lograr esas comparaciones se requiere del desarrollo de mediciones con la calidad necesaria para realizarlas, lo que plantea ciertos desafíos dependiendo del tipo de atributo del que se trate. Por ejemplo, si se tiene dos grupos de personas, para comparar su altura se puede utilizar el promedio de cada grupo. Dado que es una característica observable, se puede medir o preguntar directamente. Es decir, se puede medir utilizando un ítem. En contraste, si se quiere comparar algo más abstracto como la actitud hacia los migrantes, ya no es posible preguntar o medirla directamente. Por tanto, se requiere un instrumento que capture adecuadamente dicha complejidad, típicamente una escala compuesta de varios ítems. Como consecuencia, la comparación del promedio directo ya no es posible, dado que ciertas características de cada grupo pueden hacer que los ítems que midan esa actitud se entiendan de manera distinta. Si fuera así, sería como medir a cada grupo con una regla de medir distinta, afectando la comparación de los puntajes.

Aquellos conceptos que no pueden ser observados de manera directa, son tratados como variables latentes (Bollen, 2002). Esto quiere decir, que para medirlos se identifican una serie de aspectos observables para inferir el constructo no observable. Respecto de este tipo de variables se producen desafíos importantes, tanto para la construcción de mediciones en cuanto a una adecuada conceptualización teórica, operacionalización, y las evidencias de

validez de las mediciones, así como para la evaluación de su comparabilidad (Davidov et al., 2014; American Educational Research Association [AERA] et al., 2014; Millsap, 2011; Rammstedt et al., 2015).

En un contexto de creciente disponibilidad de datos de investigaciones sociales basados en encuestas (e.g., International Social Survey Program [ISSP], Encuesta Mundial de Valores, estudio PISA, Trends in International Mathematics and Science Study [TIMMS], etc.), tanto a nivel nacional como internacional, se presenta la posibilidad de estudiar comparativamente una amplia gama de constructos latentes tales como percepción de desigualdad social, actitudes sociales y políticas, actitudes hacia la democracia, participación política, calidad de vida, por mencionar algunos relevantes. En general estos estudios ponen a disposición de los usuarios documentación acerca de los conceptos y evidencias de validez acerca de las mediciones (Rammstedt et al., 2015), es decir que las mediciones cuentan con evidencias y teoría que respaldan sus usos (AERA et al., 2014). Esto implica que es posible usar las mediciones disponibles en estos estudios. Sin embargo, algunas investigaciones muestran posibles problemas en su utilización directa. Instrumentos descritos como unidimensionales podrían en realidad funcionar mejor al considerar su multidimensionalidad (Miranda & Castillo, 2018) o diferentes escalas o indicadores que se proponen medir el mismo concepto podrían en realidad medir cosas diferentes (Castillo et al., 2012).

La comparabilidad de las mediciones entre un número pequeño de grupos (e.g., dos países) o entre un número grande (e.g., 20 países o más), requiere que las mediciones puedan ser interpretadas de manera equivalente. La disponibilidad de evidencia acerca de la comparabilidad de las mediciones es información mucho menos disponible en la documentación de estudios basados en encuestas y es un tema de amplia discusión en la literatura reciente (Liu et al., 2017; Meuleman et al., 2022; Rutkowski & Svetina, 2014; van de Schoot et al., 2015). Uno de los principales desafíos en los estudios comparativos es lograr la invarianza estadística de las medidas, lo que permitiría una comparabilidad con sentido sustantivo (Davidov et al., 2014; Millsap & Meredith, 2007). Así, interpretar las diferencias de una medición entre dos o más grupos (e.g., niveles de autoritarismo entre personas de distintos países) implica que esa medición refleja las verdaderas diferencias entre estos grupos, más que sólo diferencias en la medición misma. Entonces, al utilizar las mediciones disponibles en estudios comparados: ¿en qué medida se cumplen las condiciones necesarias para comparar los conceptos entre países? Este artículo se propone proveer evidencia acerca de la comparabilidad de mediciones en estudios internacionales y las potenciales consecuencias para la evaluación de hipótesis. A partir de un ejemplo, las actitudes hacia la igualdad de derechos entre hombres y mujeres disponible en el Estudio Internacional de Educación Cívica y Ciudadana (Schulz et al., 2016), se propone, primero, comparar el modelo original propuesto en el estudio (unidimensional) versus un modelo alternativo (bidimensional), siguiendo discusiones previas al respecto (Castillo et

al., 2019; Miranda & Castillo, 2018). Segundo, se propone evaluar el grado de comparabilidad de las mediciones, testeando la invarianza métrica de estas.

Medición en ciencias sociales

En cualquier disciplina con aproximación cuantitativa la medición es un tema clave. Particularmente, en ciencias sociales la medición implica la definición de constructos o conceptos teóricos que especifica los dominios de su significado y la construcción de instrumentos que asignen una métrica interpretable. A grandes rasgos, se puede distinguir entre aspectos tangibles que pueden ser medidos de manera directa (e.g., edad, estatus marital, número de hijos) y aquellos aspectos intangibles que no pueden observados directamente, como actitudes, valores o creencias. Dada esta condición, estos tienden a conceptualizarse como constructos latentes. Es decir, no son observados ni medidos directamente, pero pueden ser operacionalizados a partir de indicadores indirectos que conformen un instrumento de medición. Esto implica que no se pueden medir utilizando un sólo ítem, sino que deben medirse a partir de múltiples indicadores (Beaujean, 2014). Por ejemplo, la autoeficacia es un constructo latente que no es posible de observar de manera directa. Para medirla se ha desarrollado instrumentos que identifican un listado de ítems que cuantifican los aspectos esenciales en que darían cuenta del constructo. El desafío es que los instrumentos diseñados no sólo deben medir, sino que deben hacerlo de manera válida, lo que implica seguir una serie de estándares orientados a garantizar la calidad de las mediciones.



Se observan crecientes esfuerzos para proponer estándares que permiten definir criterios de calidad para las mediciones en ciencias sociales. Estos estándares se enfocan en la acumulación de evidencia y teoría que respalda la interpretación de los puntajes de un instrumento y en la coherencia de los puntajes en el procedimiento de evaluación. Por un lado, tres instituciones del ámbito de la evaluación educacional y del desarrollo de test psicométricos (i.e., AERA, American Psychological Association [APA], y National Council on Measurement in Education [NCME]) desarrollaron esfuerzos colaborativos en esa línea, materializados en una serie de estándares publicados en su primera versión en el año 1954, hasta su última versión del año 2014 (AERA et al., 2014). Por otro lado, aunque la discusión no ha estado ausente en la Sociología, la definición de estándares tiende a ser más reciente. Por ejemplo, en 2015 el Quality Standards Working Group en el marco del German Data Forum proponen un documento que define estándares de calidad para el desarrollo, aplicación y evaluación de instrumentos de medición en investigaciones basadas en encuestas (ver: Rammstedt et al., 2015). Esta propuesta está alineada con los estándares definidos desde la educación y la psicología, dado que se centra en el proceso de construcción de instrumentos (conceptualización y operacionalización) para acumular evidencia acerca del funcionamiento de las mediciones y la correcta inferencia que se puede hacer sobre el constructo.

Múltiples técnicas estadísticas han sido desarrolladas en línea con los estándares de calidad de las mediciones, buscando mejorar el proceso de construcción de instrumentos que

permita una mejor interpretación y también comparación de los puntajes a la luz de la teoría. Uno de los enfoques más utilizados para aproximarse a la medición de conceptos no observables directamente es el que los define como variables latentes (Bollen, 2002). Este enfoque procura que los constructos hipotéticos subyacentes sean capturados adecuadamente por un conjunto de indicadores observables mediante el uso de técnicas estadísticas. Sin embargo, no son pocos los casos en que se producen dificultades, tanto de definición conceptual, como en su posterior operacionalización. En algunas ocasiones las mismas etiquetas conceptuales se refieren de manera imprecisa a constructos distintos o en algunas ocasiones constructos distintos son medidos a partir instrumentos similares. Dos ejemplos. Como bien se ha documentado, el concepto de compromiso cívico (*civic engagement*), tiene múltiples acepciones, lo que implica que no siempre se refiere al mismo objeto teórico. Es más, su medición llega a ser inconsistente (Ekman & Amnå, 2012). Por otro lado, conceptos diferentes (e.g., sofisticación política, conocimiento político, conciencia política) que tienen claras distinciones en su definición, habitualmente se miden usando el mismo set de preguntas (Miranda & Herrera-Leighton, 2021; Rapeli, 2013).

En suma, el estudio cuantitativo de fenómenos de interés para las ciencias sociales tiene múltiples desafíos que provienen de los múltiples pasos que se requieren para lograr una medición y, por tanto, una interpretación adecuada de los constructos. Más aún si se trata de hacer estudios comparativos, donde los mismos constructos y mediciones son utilizados para comparar contextos.



Comparabilidad de las mediciones: Invarianza métrica

La discusión en torno a la necesidad de contar con mediciones equivalentes para interpretar correctamente las diferencias es creciente, bajo la premisa de atribuir las diferencias observadas en una comparación a diferencias verdaderas y no a diferencias en el funcionamiento del instrumento utilizado para medirlas (Davidov, 2009; Davidov et al., 2014; Meuleman et al., 2022; Rutkowski & Svetina, 2014; Svetina & Rutkowski, 2017). Así, comparar mediciones que no cuentan con evidencia necesaria para asegurar su comparabilidad tiene una serie de consecuencias relevantes. Un estudio de simulación que analiza el impacto de comparar dos sociedades utilizando escalas no-equivalentes, muestra que las diferencias de funcionamiento del instrumento se pueden traducir en falsos efectos de interacción o en diferencias de media artificiales (Chen, 2008). En la misma dirección, un estudio reciente (Lacko et al., 2022) compara una escala de individualismo/colectivismo invariante versus la misma escala usando una aproximación tradicional (sin testear su equivalencia métrica) y muestra que se producen importantes inconsistencias entre ambas aproximaciones en los resultados de posteriores pruebas estadísticas que utilizan las escalas.

En el marco del desarrollo de técnicas para mejorar la comparabilidad de las mediciones, la más relevante es el análisis factorial confirmatorio (AFC), orientado a evaluación la dimensionalidad latente de un constructo a partir de una serie de indicadores observables. Esta técnica es considerada indispensable para proveer evidencias de validez

en ciencias sociales y del comportamiento (Brown, 2015). Sin embargo, un buen ajuste factorial confirmatorio no garantiza la comparabilidad entre grupos. Para ello se requieren técnicas adicionales que posibilitan testear el grado de comparabilidad de las mediciones entre distintos grupos. Una extensión adicional, que ha tomado creciente protagonismo en el marco de estudios que incluyen múltiples países, permite evaluar cuán generalizable es un modelo de medición entre diferentes grupos (e.g., un modelo aplicado en diferentes países). Esta evaluación de la comparabilidad se conoce técnicamente como invarianza métrica (Millsap, 2011) o equivalencia métrica (Davidov et al., 2014). En la última década se ha extendido el uso de técnicas estadísticas desarrolladas para evaluar esta equivalencia y mejorar las comparaciones. Una de estas técnicas, cuyo uso está más extendido es el AFC Multigrupo (MGCFA), la cual permite evaluar la comparabilidad de medidas a través de la estimación de un modelo de medición de manera secuencial. Cada paso sucesivo de estimación agrega restricciones paramétricas crecientes que representan distintos niveles de comparabilidad (invarianza).

El primer nivel de invarianza se conoce como *invarianza configural*. El modelo configural asume que el constructo latente es medido por los mismos indicadores en todos los grupos. Es decir, si se pretende comparar mediciones entre países, se debe especificar el mismo modelo de medición para cada uno de ellos. Este es el modelo de referencia o modelo base, que evalúa el ajuste estadístico de la misma configuración de indicadores que capturan el concepto latente



para todos los grupos. Este nivel no garantiza ninguna comparabilidad entre grupos (Beaujean, 2014), pero da cuenta de la adecuación basal del modelo de medición. Esto implica que, si el ajuste del modelo configurar es pobre, difícilmente habrá posibilidad de que además sea un modelo comparable entre grupos.

El segundo nivel le conoce como *invarianza métrica*. El modelo métrico, también conocido como invarianza débil, restringe las cargas factoriales para que sea la misma en todos los grupos. Esta restricción considera no solo la misma configuración, sino define que cada indicador tiene la misma carga factorial en todos los grupos considerados. Así, un cambio estadísticamente relevante en el ajuste, indicaría que el modelo de medición no funciona del mismo modo para los distintos grupos, al contrario un cambio estadístico mínimo en el ajuste indica que el modelo funciona con cargas factoriales equivalentes para todos los grupos. Este nivel de invarianza se considera el primer nivel de comparabilidad propiamente tal y sería la condición mínima para hacer comparaciones de las variables latentes en términos de sus relaciones con otras covariables (Beaujean, 2014; Davidov et al., 2014).

El tercer nivel se le conoce como *invarianza escalar* o invarianza fuerte. En esta estimación sucesiva, además de la misma configuración (modelo configural) con la restricción de las cargas factoriales (modelo métrico), se agrega la restricción de igualdad de los interceptos para todos los grupos en el caso de variables continuas y de umbrales en todos los grupos, en el caso de modelos con variables categóricas (Liu et al., 2017; Millsap & Yun-Tein, 2004; Rutkowski & Svetina, 2014). La

comparación del ajuste del modelo escalar con el modelo métrico permite evaluar si este nivel de comparabilidad se logra (o no). Un cambio estadístico sustantivo en el ajuste global indica que el modelo de medición no funciona del mismo modo para los distintos grupos, es decir, los interceptos del modelo no serían comparables. Al contrario, un cambio estadístico mínimo en el ajuste indica que el modelo funciona con interceptos (o umbrales) equivalentes para todos los grupos. Lograr este nivel de invarianza permite comparaciones de los promedios latentes entre grupos así como la comparaciones de la relación de variables latentes con otras covariables. En ese sentido, este nivel de invarianza permite la comparación promedios de un constructo entre los grupos, y/o el uso de modelos estadísticos, como regresiones, que comparan patrones relacionales entre grupos.

Finalmente, hay un cuarto nivel llamado invarianza *estricta*. Este nivel agrega la restricción de la variación de los errores residuales de los indicadores entre grupos, lo que aumenta la comparabilidad de las escalas latentes. Sin embargo, dado que la invarianza escalar es suficiente para realizar comparaciones estadísticas entre medias de grupos y patrones de covariables, a menudo no se estima el nivel de invarianza estricta (Beaujean, 2014; Davidov et al., 2014).

Este estudio

Una de las bases de datos disponibles para hacer estudios comparativos es el International Civic and Citizenship Study (ICCS) (International Association for the Evaluation of Educational Achievement [IEA]- ICCS, 2016), que cuenta con



dos aplicaciones en las que Chile ha participado: 2009 y 2016. Esta investigación tiene como finalidad analizar la manera en que jóvenes de octavo grado se encuentran preparados para asumir su rol como ciudadanos. En esa dirección se proponen medir múltiples aspectos enmarcados en conocimientos, actitudes valores y expectativas en torno al ejercicio de la ciudadanía (Schulz et al., 2016, 2018).

El estudio ICCS hace un esfuerzo importante por cuidar el proceso de construcción de los instrumentos utilizados y así asegurar la comparabilidad de las mediciones. En ese sentido, proveen amplia documentación sobre los procedimientos de construcción de instrumentos, evidencias de validez, y disponibilidad abierta de los datos. A pesar de ello, es posible que los instrumentos desarrollados presenten dificultades en su comparabilidad entre países, dada la dificultad de lograrla y la cantidad de escalas utilizadas en el estudio.

Nuestro artículo pone foco en una de estas escalas presentes en ICCS 2016: el instrumento que mide actitudes hacia la igualdad de derechos entre hombres y mujeres. En el marco de los ideales democráticos, la disposición positiva hacia la igualdad de derechos de todos los grupos es esperable, por lo tanto, esta actitud particular es definida como una actitud hacia un principio cívico. El concepto se concibe como unidimensional, es decir, no tendría dimensiones o facetas que lo hagan más complejo. Además,

consistente con esta concepción se propone medir estas actitudes a partir de una escala de seis ítems orientados a capturar aspectos que darían cuenta de esta actitud, utilizando el marco de variables latentes mencionado previamente. Sin embargo, investigaciones previas cuestionan su unidimensionalidad (Castillo et al., 2019; Miranda & Castillo, 2018). Por un lado, se propone que las actitudes hacia la igualdad de derechos entre hombres y mujeres ha sido conceptualizada previamente como un concepto multidimensional. Es decir, habría ámbitos de la vida en que esta actitud puede operar de manera diferenciada. Por ejemplo, se distingue la disposición general hacia la igualdad, de la disposición hacia la igualdad en la vida privada o en el espacio público. Por otro lado, la evidencia empírica tiende a mostrar que en la versión del año 2009, un modelo de dos dimensiones funciona mejor que el modelo unidimensional originalmente propuesto.

El presente estudio se propone evaluar la adecuación métrica del concepto de actitudes hacia la igualdad de derechos contrastando un modelo unidimensional, disponible en la documentación original del estudio comparado, con un modelo bidimensional, usando los datos de la nueva versión ICCS del año 2016. Además, se propone evaluar los niveles de comparabilidad de las mediciones por medio del MGCFA.

MÉTODO

Datos

Se usó la base de datos del ICCS (2016) que contempla 94,000 estudiantes de octavo grado en 3,800 escuelas de 24 países. Los datos de estudiantes se complementaron con los de más de 37,000 profesores al interior de las escuelas y con datos contextuales recogidos de los directores de las escuelas y de los centros de investigación nacionales. El cuestionario europeo de estudiantes recogió datos de 53,000 estudiantes en 14 países. El cuestionario de estudiantes latinoamericanos recogió datos de más de 25,000 estudiantes de cinco países (Köhler, et al., 2018).

La estrategia de muestreo fue aleatoria, estratificada y multi-etápico. En cada país, se tomaron muestras de 150 escuelas aproximadamente, dentro de las cuales se incluyen estudiantes de un aula de clase y 15 profesores. La muestra analítica final, considera los datos válidos con información completa correspondiente a 82,285 estudiantes. La Tabla 1 muestra la distribución por país de las escuelas y la muestra de estudiantes.

Variables

Las variables del estudio corresponden a los indicadores de ICCS 2016 que conforman una escala de apoyo a la igualdad de género de parte de los estudiantes. La escala original, consideró los seis indicadores que son entendidos como parte de un único factor latente que se refiere a la actitud de los

hacia la igualdad de derechos o capacidades entre hombres y mujeres.

Tabla 1

Muestra de Países Participantes en ICCS

País	Total escuelas	Total estudiantes
Alemania	59	1,196
Bélgica	162	2,566
Bulgaria	147	2,602
Chile	178	4,508
Colombia	150	4,829
Corea del Sur	93	2,447
Croacia	175	3,661
Dinamarca	183	5,103
Eslovenia	145	2,605
Estonia	164	2,613
Finlandia	179	2,797
Italia	170	3,203
Letonia	147	2,733
Lituania	181	3,162
Malta	47	3,040
México	211	4,861
Noruega	148	5,265
Países Bajos	123	2,485
Perú	206	4,576
RAE de Hong Kong (China)	91	2,347
República Dominicana	141	2,949
Rusia	352	6,484
Suecia	154	2,454
Taiwán	141	3,799



Los indicadores A, B, y E en la Tabla 2 corresponden al nivel de acuerdo frente a afirmaciones en torno a derechos, participación, y oportunidades políticas. Segundo, los indicadores C, D, y F en la Tabla 2, corresponden al nivel de acuerdo de los estudiantes en torno a temas relacionados con la superioridad/inferioridad entre hombres y mujeres en la sociedad. En suma, un mayor puntaje en esta escala indica una actitud

más favorable hacia la igualdad de derechos entre hombres y mujeres. Aquí se evalúa la uni o bidimensionalidad de la escala para evaluar la calidad de la medición y las consecuencias que esto puede tener para la investigación en ciencias sociales. En la Tabla 3 se pueden observar los estadísticos descriptivos generales de las variables del estudio.

Tabla 2

Indicadores de Apoyo de los Estudiantes a la Igualdad de Género

- a. Los hombres y las mujeres deben tener igualdad de oportunidades para participar en el gobierno (R)
- b. Los hombres y las mujeres deben tener los mismos derechos en todos los sentidos (R)
- c. Las mujeres deben permanecer fuera de la política.
- d. Cuando no hay muchos puestos de trabajo disponibles, los hombres deben tener más derecho a un trabajo que las mujeres.
- e. Los hombres y las mujeres deben obtener la igualdad de salario cuando están haciendo los mismos trabajos (R)
- f. Los hombres están mejor calificados para ser líderes políticos que las mujeres.

Nota. La escala de respuesta utilizada corresponde a (1) Muy de acuerdo; (2) De acuerdo; (3) En desacuerdo; (4) Totalmente en desacuerdo. (R) Ítems invertidos

Estrategias de análisis

El análisis consiste en cuatro momentos. Primero, el modelo de medición original de una dimensión fue evaluado usando AFC para cada país por separado. Luego, el modelo original fue contrastado con un modelo de medición bidimensional, también empleando AFC. Para establecer y comparar el nivel de ajuste del modelo de medición original y bidimensional se empleó el test de chi-cuadrado, el cual constituye una primera aproximación al ajuste del modelo de medición. Debido a la sensibilidad al tamaño muestral de la prueba chi-cuadrado, se

emplean de manera complementaria el Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) y el Root Mean Square Error Approximation (RMSEA) (Brown, 2015). Una bondad de ajuste aceptable para el CFI y TLI es de valores cercanos a .95 o mayores, sin embargo, valores en el rango de .90 y .95 también se consideran aceptables. Adicionalmente, un RMSEA con valores de .06 o menores fueron considerados adecuados.

Un tercer momento es el análisis de invarianza métrica multigrupos sobre el modelo de medición



propuesto de dos dimensiones, con el objetivo de conocer el nivel de estabilidad de la medición entre distintos contextos culturales, lo cual constituye un interés en ascenso en la literatura sociológica (Davidov et al., 2014). La literatura generalmente sugiere que este modelamiento se debe realizar sobre la base de cuatro niveles de restricción progresiva (Cheung & Rensvold, 2002; Milfont & Fischer, 2010): (1) Configuración: Se estima el Modelo solo de indicando la estructura de este, razón por lo cual todos los parámetros se estiman libremente. (2) Métrica o Débil: Posteriormente, en el modelo se restringen las cargas factoriales de cada ítem, haciéndolas idénticas entre los grupos. (3) Escalar o Fuerte: Luego, se estima el modelo restringiendo las cargas factoriales y los interceptos de cada ítem, haciéndolos idénticos entre los grupos. (4) Estricta: Finalmente, se estima un modelo donde se restringen las cargas factoriales, los interceptos y las varianzas de error para cada ítem, haciéndolos idénticos entre los grupos.

La literatura sobre invarianza métrica sugiere dos aproximaciones complementarias para el testeo de invarianza. En nuestro caso, usaremos la comparación en base al ajuste incremental de los estadísticos de ajuste (Cheung & Rensvold, 2002; Milfont & Fischer, 2010) y el test ANOVA para

comparación de medias en modelos anidados (Newsom, 2015).

Adicionalmente, y tomando en consideración las aproximaciones críticas frente a los criterios convencionales en el análisis de invarianza, la literatura ha sugerido adoptar un enfoque relativo para la evaluación de la invarianza métrica en múltiples grupos, tales como son las comparaciones internacionales, donde las unidades o clusters corresponden a países (Svetina & Rutkowski, 2017). Esta aproximación propone abordar el ajuste relativo de los modelos anidados, principalmente a través del análisis de la significancia estadística del chi-cuadrado y observando la magnitud del cambio de los estadísticos CFI y RMSEA.

Respecto de la invarianza débil o métrica, se consideró aceptables cambios mayores o iguales a $-.004$ en el CFI y cambios menores o iguales a $.05$ en el RMSEA. Para la invarianza fuerte o escalar, se consideró adecuados cambios mayores o iguales a $-.01$ en el CFI y cambios menores o iguales a $.01$ en el RMSEA (Svetina & Rutkowski, 2017). En este sentido, la equivalencia escalar permite establecer comparaciones válidas entre las medias latentes de los grupos, lo cual cobra relevancia sustantiva en los análisis multinivel, en donde se asume que las medias entre países o grupos son comparables (Davidov et al., 2014)

Tabla 3

Estadísticos Descriptivos Para Variables de Estudiantes

Ítem	Categorías	Frec. (%)	Gráfico
Los hombres y las mujeres deben tener igualdad de oportunidades para participar en el gobierno (R).	MA	577 (0.7%)	
	A	1,870 (2.3%)	
	D	17,239 (21.0%)	
	MD	62,599 (76.1%)	
Los hombres y las mujeres deben tener los mismos derechos en todos los sentidos (R).	MA	690 (0.8%)	
	A	3,904 (4.7%)	
	D	17,612 (21.4%)	
	MD	60,079 (73.0%)	
Las mujeres deben permanecer fuera de la política.	M	4,967 (6.0%)	
	A	7,820 (9.5%)	
	D	23,936 (29.1%)	
Cuando no hay muchos puestos de trabajo disponibles, los hombres deben tener más derecho a un trabajo que las mujeres.	MD	45,562 (55.4%)	
	MA	7,028 (8.5%)	
	A	10,188 (12.4%)	
	D	23,684 (28.8%)	
Los hombres y las mujeres deben obtener la igualdad de salario cuando están haciendo los mismos trabajos (R).	MD	41,385 (50.3%)	
	MA	1,530 (1.9%)	
	A	4,466 (5.4%)	
	D	17,451 (21.2%)	
Los hombres están mejor calificados para ser líderes políticos que las mujeres.	MD	58,838 (71.5%)	
	MA	6,810 (8.3%)	
	A	11,578 (14.1%)	
	D	26,097 (31.7%)	
Nivel educacional (Padres)	<Terciaria.	37,800 (45.9%)	
	>Terciaria	50,242(61.1%)	
Apoyo de los estudiantes a la igualdad de género (original)	19 valores distintos. Media (DE): 51.6 (9.8). min ≤ mediana ≤ max: 16.3 ≤ 48.7 ≤ 63.9 RI (CV) = 21.3 (0.2)		
Puntaje factorial: Igualdad de derechos políticos	Variable continua. Media (DE): 88.2 (16.2). min ≤ mediana ≤ max: 0 ≤ 97 ≤ 100 RI (CV) = 19.5 (0.2)		
Puntaje factorial: Participar en el espacio público	Variable continua. Media (DE): 76.5 (23.9), min ≤ mediana ≤ max: 0 ≤ 80.2 ≤ 100 RI (CV) = 38.8 (0.3)		

Nota. MA=muy de acuerdo. A= En acuerdo. D= en desacuerdo. MD= muy en desacuerdo. DE = desviación estándar. RI = rango intercuartílico. CV = coeficiente de variación. El número de casos válidos en cada uno de los ítems de esta tabla correspondió a 82,285.



RESULTADOS

Análisis descriptivo

La Tabla 4 muestra las correlaciones policóricas de los indicadores de actitudes hacia la igualdad de género de estudiantes de octavo grado para la muestra completa ($n = 82,285$). A modo general los resultados de las correlaciones muestran, primero, que asumir unidimensionalidad del constructo respecto a las actitudes hacia la igualdad de género es puesto en cuestión al observar que el tamaño de las correlaciones difiere entre los indicadores C, D, y F, en contraste con los indicadores A, B y E, formándose dos grupos según el tamaño de las correlaciones. Por un lado, vemos que los indicadores C, D y F, referentes a las afirmaciones

sobre la participación de las mujeres en el espacio público muestran correlaciones altas entre sí. Por ejemplo, el ítem en el cual se afirma que “Las mujeres deben permanecer fuera de la política” posee una correlación positiva y estadísticamente significativa con la afirmación respecto a que el “Trabajo debe ser prioridad para los hombres” ($r = .71, p < .01$) y con que “los hombres están mejor calificados para ser líderes políticos” ($r = .68, p < .01$). Por el otro lado, los indicadores A, B y E, agrupados en torno a la idea de la igualdad de derechos políticos entre hombres y mujeres, muestran correlaciones medias y altas.

Tabla 4

Matriz de Correlaciones Policóricas

Ítems de apoyo de a la igualdad de género	(A)	(B)	(C)	(D)	(E)
A. Hombres/Mujeres igualdad en gobierno (r)	-				
B. Hombres/Mujeres mismos derechos (r)	.78	-			
C. Mujeres permanecen fuera de la política	.47	.46	-		
D. Trabajo, prioridad para hombres	.46	.47	.71	-	
E. Hombres/mujeres igualdad salarial (r)	.61	.62	.39	.43	-
F. Hombres mejores lideres políticos	.47	.47	.68	.75	.44

Nota. Todas las correlaciones son significativas a $p < .001$.

Por ejemplo, el indicador en el cual se afirma que “Los hombres y las mujeres deben tener los mismos derechos en todos los sentidos” posee una correlación positiva y significativa con que hombres y mujeres tienen igual derecho a participar de un gobierno ($r = .78; p < .01$), así como también

con la afirmación sobre el derecho a la igualdad salarial ($r = .62; p < .01$). Finalmente, todas las correlaciones entre los indicadores de estos dos grupos son positivas, pero con una intensidad menor.

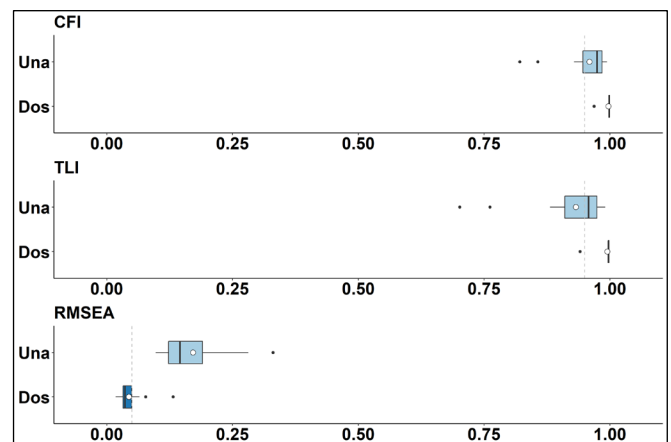
Análisis factorial confirmatorio

El modelo de medición original propuesto por ICCS establece una variable latente unidimensional que explica los seis indicadores. No obstante, el análisis de la matriz de correlaciones indica que asumir un constructo unidimensional podría quedar en duda a partir del patrón de correlaciones previamente presentado. Considerando estos antecedentes, consistentes con investigaciones previas (Castillo et al., 2019 Miranda & Castillo, 2018) se estimaron dos modelos, el primero unidimensional y el segundo bidimensional.

Siguiendo los criterios convencionales para evaluar la bondad de ajuste del modelo de medición, se muestra que el modelo de una dimensión no posee un ajuste aceptable ($\chi^2(gf) = 25258.04$ (216), CFI = .97, TLI = .95, RMSEA = .18), particularmente en cuanto al indicador de parsimonia RMSEA. En contraste con el modelo de dos dimensiones que muestra una mejora substantiva en el ajuste ($\chi^2(gf) = 1915.63$ (192), CFI = .99, TLI = .99, RMSEA = .05). Estos resultados sugieren que el modelo bidimensional, en contraste con el modelo de una dimensión, posee un mejor ajuste y constituye una medición más precisa del constructo propuesto. Consistente con los resultados para la muestra completa, la Figura 2 muestra la distribución de los estadísticos de bondad de ajuste para el análisis factorial confirmatorio para cada uno de los 24 países de la muestra (ver detalles en Anexo). Respecto del modelo de unidimensional, un 75% de los países ($n = 18$) obtiene un ajuste aceptable de acuerdo con CFI. Un 58.3% ($n = 14$) de los países obtiene un TLI con valores considerados aceptables.

Sin embargo, en ningún país el modelo de medición obtuvo valores aceptables para el RMSEA. A nivel general, el modelo de dos dimensiones es una mejora substantiva en todos los países. En todos los países ($n = 24$) el modelo bidimensional obtuvo un CFI aceptable; en un 95.8% de los países ($n = 23$) obtuvo un TLI aceptable y en un 83.3% de los países ($n = 20$) obtuvo un RMSEA igual o menor a .06. En suma, considerando conjuntamente los estadísticos de ajuste, más de la mitad de la muestra de países mejoró su bondad de ajuste en el modelo de dos dimensiones, en contraste con el original de una dimensión.

Figura 2. *Ajuste del Modelo de Medición de Igualdad de Género Según Número de Dimensiones*



Invarianza métrica

Los resultados respecto al número de dimensiones del modelo de medición para las actitudes hacia la igualdad de género de estudiantes de octavo grado son claros en que la mejor especificación es la de dos dimensiones. Posteriormente, se realizó un testeó de



invarianza métrica multigrupo para el modelo bidimensional, donde cada grupo corresponde a un país, es decir, se puso a prueba la estabilidad del modelo de medición en diferentes contextos culturales. La Tabla 5 muestra los resultados para la aproximación incremental de invarianza y del test ANOVA para comparación de ajustes entre los niveles de invarianza. En este test se esperaba no rechazar la hipótesis nula, es decir, que las diferencias en el ajuste no fueran estadísticamente significativas. El ajuste basado en la prueba chi-

cuadrado muestra una disminución sustantiva entre los niveles de invarianza. Sin embargo, los resultados de la prueba ANOVA para la diferencia del estadístico chi-cuadrado muestran que no se cumple para el nivel de invarianza débil o métrica al contrastarlo con el modelo configural, mostrando diferencias estadísticamente significativas entre los ajustes ($\Delta\chi^2(\Delta gl) = 757.77 (92), p < .01$). Adicionalmente, el contraste de hipótesis para el nivel de invarianza fuerte y estricta no ofrece evidencia en favor de la invarianza de intercepto y varianza residual.

Tabla 5

Invarianza Multigrupo Para Modelo Bidimensional de Igualdad de Género

Modelo	$\chi^2(gl)$	CFI	RMSEA (90 IC)	$\Delta\chi^2(\Delta gl)$	ΔCFI	$\Delta RMSEA$
Configural	1915.6 (192)	.998	.051 (.049-.053)			
Débil	2673.4 (284)	.997	.050 (.048-.051)	757.8 (92)*	-0.001	-0.002
Fuerte	9743.3 (514)	.988	.072 (.071-.074)	7069.9 (230)*	-0.009	0.023
Estricta	18145.8 (652)	.978	.088 (.087-.090)	8402.5 (138)*	-0.011	0.016

Nota. *** $p < .001$. Modelos son calculados usando el estimador DWLS para indicadores categóricos ordinales ($n = 82,285$). IC = Intervalo de confianza. gl = grados de libertad.

En contraste al enfoque convencional detallado anteriormente, la evaluación de la invarianza desde una aproximación del cambio relativo en los estadísticos CFI y RMSEA, muestra que el modelo de invarianza débil posee un cambio en el CFI de -.001 respecto al modelo configural, como también una diferencia de -.002 en el RMSEA. Estos resultados indican que las restricciones del modelo de invarianza métrica no deterioran sustantivamente el ajuste del modelo, por lo tanto, es posible asumir invarianza a nivel de las cargas factoriales entre los grupos. Luego, el ajuste del modelo para invarianza

fuerte muestra una diferencia de -.009 en el CFI, mientras que el cambio en el estadístico RMSEA es de .023, ambos por debajo de los criterios establecidos (Svetina & Rutkowski, 2017). Estos resultados sugieren que el ajuste del modelo para invarianza fuerte, en el cual se restringen cargas factoriales e interceptos, es satisfactoria, lo cual permite asumir que las medias latentes de cada país en el modelo de medición son equivalentes y, por tanto, es posible realizar comparaciones transculturales sustantivas (Davidov et al., 2014; Svetina & Rutkowski, 2017).



CONCLUSIONES

En un contexto de creciente disponibilidad de datos de investigaciones sociales basados en encuestas se presenta la posibilidad de estudiar comparativamente una amplia gama de constructos latentes. Sin embargo, algunas investigaciones muestran posibles problemas en su utilización directa. En ese marco este estudio abordó la siguiente pregunta de investigación: ¿en qué medida se cumplen las condiciones necesarias para comparar los conceptos entre países? A partir de la evaluación del concepto de actitudes hacia la igualdad de derechos entre hombres y mujeres y la escala para su medición (ver: ICCS, 2016), se comparó la estimación del modelo original unidimensional versus un modelo bidimensional. La principal conclusión es el modelo bidimensional es una mejor representación del constructo teórico propuesto. Por un lado, en el modelamiento de los datos agregados, usándolos 24 países incluidos en el estudio, se observa que el modelo unidimensional presenta peores indicadores de ajuste, mientras que el modelo bidimensional logra mejoras sustanciales. Por otro lado, la evaluación de la invarianza métrica del modelo bidimensional muestra tanto invarianza métrica como escalar. Esto quiere decir que el instrumento especificado en sus dos dimensiones posee una comprensión similar por parte de los participantes en los distintos países de la muestra. La implicancia principal es que las dimensiones de las actitudes hacia la igualdad de derechos entre hombres y mujeres logran importantes grados de equivalencia para su interpretación. Con los resultados obtenidos, se podría comparar los niveles de cada dimensión (e.g., comparación de medias latentes) y/o la

relación de estas dimensiones con otras variables (e.g., comparación de modelos de regresión).

Los resultados de este estudio muestran la importancia de evaluar activamente la construcción de instrumentos disponibles en estudios secundarios al momento de comparar países. Se provee evidencia acerca de la comparabilidad de mediciones en estudios internacionales y las potenciales consecuencias para la evaluación de hipótesis relacionales y la comparabilidad de los niveles entre países. Como se muestra, el modelo de medición que se especifique podría alterar las conclusiones, por lo que la generación de evidencias acerca de la comparabilidad de las mediciones entre grupos se torna en un set de herramientas indispensables para el estudio fenómenos comparados. Al igual que en estudios previos (Lacko et al., 2022) se destaca la necesidad del uso de técnicas que evalúen la comparabilidad de las mediciones entre grupos (en este caso países) para mejorar la interpretación sustantiva de las mediciones.

Lograr invarianza de las mediciones es un desafío importante para el estudio comparado y presenta múltiples desafíos que concitan amplia discusión académica en torno a las dificultades que se presentan. La literatura discute las características de las mediciones que pueden condicionar la comparabilidad (Svetina & Rutkowski, 2017), entre las que se encuentran: la complejidad de las mediciones (modelos unidimensionales versus modelos con múltiples dimensiones), el número de indicadores disponibles en el estudio, los niveles de medición de los indicadores (p.e., indicadores continuos, ordinales o dicotómicos),



el número de grupos involucrados en la comparación, por mencionar los temas más salientes. En línea con las dificultades y su creciente uso se han desarrollado aproximaciones más laxas (Flake & McCoach, 2018; Lomazzi, 2018; Magraw-Mickelson et al., 2020; Munck et al., 2018) que hacen uso del desarrollo de herramientas estadísticas recientes. Por ejemplo, el método de optimización de invarianza aproximada *alignment optimization* o modelos de bayesianos de invarianza métrica (Davidov et al., 2015; van de Schoot et al., 2015). Adicionalmente,

se propone el uso de modelos que permitan explicar factores a nivel de país que puedan dar cuenta del grado observado de invarianza, no con afán de lograrla, pero sí de explicarla (Davidov et al., 2018). Como se aprecia, el tópico de la comparabilidad está aún en desarrollo y su discusión es contingente (Meuleman, et al., 2022; Welzel et al., 2021). La discusión sobre nuevas técnicas para evaluar invarianza sigue abierta ante la necesidad de lograr comparar lo comparable en estudios internacionales.

REFERENCIAS

- American Educational Research Association., American Psychological Association., & National Council on Measurement in Education. (2014). *The standards for educational and psychological testing*. American Educational Research Association.
- Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1), 605–634. <https://doi.org/10.1146/annurev.psych.53.100901.135239>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2da ed.). The Guilford Press.
- Castillo, J. C., Miranda, D., & Bonilla, A. (2019). *Measuring attitudes towards gender equal rights in international assessments*. SocArXiv. <https://doi.org/10.31235/osf.io/t87se>
- Castillo, J. C., Miranda, D., & Carrasco, D. (2012). Percepción de desigualdad económica en Chile: Medición, diferencias y determinantes. *Psyche (Santiago)*, 21(1), 99–114. <http://doi.org/10.4067/S0718-22282012000100007>
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95(5), 1005–1018. <https://doi.org/10.1037/a0013193>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255. http://doi.org/10.1207/S15328007SEM0902_5
- Davidov, E. (2009). Measurement equivalence of nationalism and constructive patriotism in the ISSP: 34 countries in a comparative perspective. *Political Analysis*, 17(1), 64–82. <https://doi.org/10.1093/pan/mpn014>
- Davidov, E., Cieciuch, J., Meuleman, B., Schmidt, P., Algesheimer, R., & Hausherr, M. (2015). The



- comparability of measurements of attitudes toward immigration in the European social survey: Exact versus approximate measurement equivalence. *Public Opinion Quarterly*, 79(S1), 244–266. <https://doi.org/10.1093/poq/nfv008>
- Davidov, E., Dülmer, H., Cieciuch, J., Kuntz, A., Seddig, D., & Schmidt, P. (2018). Explaining measurement nonequivalence using multilevel structural equation modeling: The case of attitudes toward citizenship rights. *Sociological Methods & Research*, 47(4), 729–760. <https://doi.org/10.1177/0049124116672678>
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40(1), 55–75. <http://doi.org/10.1146/annurev-soc-071913-043137>
- Ekman, J., & Amnå, E. (2012). Political participation and civic engagement: Towards a new typology. *Human Affairs*, 22(3), 283–300. <https://doi.org/10.2478/s13374-012-0024-1>
- Flake, J. K., & McCoach, D. B. (2018). An investigation of the alignment method with polytomous indicators under conditions of partial measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(1), 56–70. <https://doi.org/10.1080/10705511.2017.1374187>
- International Association for the Evaluation of Educational Achievement., & International Civic and Citizenship Education Study. (2016). *Inter-university consortium for political and social research*. <https://doi.org/10.3886/ICPSR37147.v1>
- Köhler, H., Weber, S., Brese, F., Schulz, W., & Carstens, R. (Eds.). (2018). *ICCS 2016 user guide for the international database*. <https://www.iea.nl/data-tools/repository/iccs>
- Lacko, D., Čeněk, J., Točík, J., Avsec, A., Dorđević, V., Genc, A., Haka, F., Šakotić-Kurbalija, J., Mohorić, T., Neziri, I., & Subotić, S. (2022). The necessity of testing measurement invariance in cross-cultural research: Potential bias in cross-cultural comparisons with individualism–collectivism self-report scales. *Cross-Cultural Research*, 56(2–3), 228–267. <https://doi.org/10.1177/106939712111068971>
- Liu, Y., Millsap, R. E., West, S. G., Tein, J.-Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22(3), 486–506. <https://doi.org/10.1037/met0000075>
- Lomazzi, V. (2018). Using alignment optimization to test the measurement invariance of gender role attitudes in 59 countries. *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology (MDA)*, 12(1), 77–103. <https://doi.org/10.12758/mda.2017.09>
- Magraw-Mickelson, Z., Carrillo, A. H., Weerabangsa, M. M., Owuamalam, C., & Gollwitzer, M. (2020). *Comparing classic and novel approaches to measurement invariance*. PsyArXiv Preprints. <https://doi.org/10.31234/osf.io/pz8u9>
- Meuleman, B., Žóltak, T., Pokropek, A., Davidov, E., Muthén, B., Oberski, D. L., Billiet, J., & Schmidt, P. (2022). Why measurement invariance is important in comparative research. A response to Welzel et al. (2021). *Sociological Methods*



- & *Research*. Publicación anticipada en línea. <https://doi.org/10.1177/00491241221091755>
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–121. <https://doi.org/10.21500/20112084.857>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.
- Millsap, R. E., & Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. En R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 131–152). Lawrence Erlbaum Associates Publishers.
- Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, 39(3), 479–515. https://doi.org/10.1207/S15327906MBR3903_4
- Miranda, D., & Castillo, J. C. (2018). Measurement model and invariance testing of scales measuring egalitarian values in ICCS 2009. En A. Sandoval-Hernández, M. M. Isac, & D. Miranda (Eds.), *Teaching tolerance in a globalized world* (Vol. 4, pp. 19–31). Springer Cham.
- Miranda, D., & Herrera-Leighton, A. (2021). Conocimiento cívico: Concepto, antecedentes y consecuencias. En C. Villalobos, M. J. Morel, & E. Treviño (Eds.), *Ciudadanías, educación y juventudes: Investigaciones y debates para el Chile del futuro* (pp. 287–326). Ediciones UC.
- Munck, I., Barber, C., & Torney-Purta, J. (2018). Measurement invariance in comparing attitudes toward immigrants among youth across Europe in 1999 and 2009: The alignment method applied to IEA CIVED and ICCS. *Sociological Methods & Research*, 47(4), 687–728. <http://doi.org/10.1177/0049124117729691>
- Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge.
- Rammstedt, B., Beierlein, C., Brähler, E., Eid, M., Hartig, J., Kerstig, M., Liebig, S., Lukas, J., Mayer, A.-K., Menold, N., Schupp, J., & Weichselgartner, E. (2015). *Quality standards for the development, application, and evaluation of measurement instruments in social science survey research* [Documento de trabajo de RatSWD No. 245]. Rat für sozial-und Wirtschaftsdaten. <http://www.econstor.eu/handle/10419/107203>
- Rapeli, L. (2013). *The conception of citizen knowledge in democratic theory*. Springer.
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31–57. <https://doi.org/10.1177/0013164413498257>
- Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). *IEA international civic and citizenship education study 2016 assessment framework*. Springer. <http://springer.com/gp/book/9783319393568>
- Schulz, W., Carstens, R., Losito, B., & Fraillon, J. (2018). *International civic and citizenship education study 2016. Technical report*. International Association for the Evaluation of Educational Achievement (IEA).



Svetina, D., & Rutkowski, L. (2017). Multidimensional measurement invariance in an international context: Fit measure performance with many groups. *Journal of Cross-Cultural Psychology*, 48(7), 991–1008. <https://doi.org/10.1177/0022022117717028>

van de Schoot, R., Schmidt, P., de Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, 6, Artículo 1064. <https://doi.org/10.3389/fpsyg.2015.01064>

Welzel, C., Brunkert, L., Kruse, S., & Inglehart, R. F. (2021). Non-invariance? An overstated problem with misconceived causes: *Sociological Methods & Research*. Publicación anticipada en línea. <https://doi.org/10.1177/00491241211995521>

Manuscrito recibido: 06-mayo-2022

Manuscrito aceptado: 13-junio-2022



ANEXO

Tabla A1

Resumen de Ajuste Según Países Para Modelo de Medición de Igualdad de Género (dos Dimensiones)

País	χ^2	n	CFI	TLI	RMSEA
Letonia	14.95	2,733	1.00	.999	.018
Alemania	15.52	1,196	.999	.999	.028
Colombia	25.30	4,829	.999	.999	.021
Italia	29.55	3,203	.999	.998	.029
Perú	33.82	4,576	.999	.998	.027
Rep. Dominicana	33.94	2,949	.996	.992	.033
Bulgaria	34.41	2,602	.998	.996	.036
Eslovenia	35.36	2,605	.999	.998	.036
Corea del Sur	36.76	2,447	.999	.998	.038
Croacia	38.51	3,661	.999	.998	.032
Bélgica	39.17	2,566	.998	.996	.039
Estonia	39.32	2,613	.998	.997	.039
Malta	40.56	3,040	.998	.997	.037
Taiwán	59.45	3,799	.999	.998	.041
Noruega	61.29	5,265	.999	.999	.036
Países Bajos	62.57	2,485	.998	.996	.052
Rusia	65.69	6,484	.998	.996	.033
Lituania	70.39	3,162	.997	.995	.050
Suecia	85.91	2,454	.998	.996	.063
Dinamarca	86.93	5,103	.999	.998	.044
Chile	96.27	4,508	.999	.997	.049
Finlandia	102.43	2,797	.998	.996	.065
RAE de Hong Kong (China)	121.01	2,347	.998	.996	.078
México	686.53	4,861	.968	.941	.132

Nota. En todos los análisis los grados de libertad fueron 8.