

Invarianza de medida: comparación de la calidad del método de alineamiento y el Análisis Factorial Confirmatorio Multigrupo bajo diferentes condiciones

María Solar Sastre

Máster en Metodología de las Ciencias del Comportamiento y de la Salud



MÁSTERES
DE LA UAM
2018 - 2019

Facultad de Psicología



**Invarianza de medida: comparación de la calidad del método de
alineamiento y el Análisis Factorial Confirmatorio Multigrupo bajo
diferentes condiciones**

**Measurement invariance: a performance comparison between the alignment
method and the Multigroup Factor Analysis under several conditions**

Alumna: María Solar Sastre

Tutor: Jesús M^a Alvarado Izquierdo

Trabajo de Fin de Máster

Máster en Metodología de las Ciencias del Comportamiento y de la Salud

Madrid, Septiembre de 2019

Índice

| | |
|---|----|
| RESUMEN | 3 |
| INTRODUCCIÓN | 4 |
| ¿Qué es la invarianza de medida y por qué es importante?..... | 4 |
| Análisis Factorial Confirmatorio Multigrupo (AFC-MG) | 5 |
| Problemas del AFC-MG. Invarianza parcial..... | 7 |
| Alternativa: invarianza aproximada. El alineamiento..... | 9 |
| Aplicaciones del alineamiento..... | 12 |
| Estudios de simulación. Ventajas e inconvenientes..... | 14 |
| Objetivo e hipótesis..... | 17 |
| MÉTODO | 19 |
| RESULTADOS | 25 |
| Sesgo de las estimaciones..... | 25 |
| Potencia del método..... | 35 |
| DISCUSIÓN Y CONCLUSIONES | 38 |
| REFERENCIAS | 46 |
| APÉNDICES | 56 |
| Apéndice A..... | 56 |
| Apéndice B..... | 69 |
| Apéndice C..... | 73 |

Resumen

Ante los problemas que puede presentar el Análisis Factorial Multigrupo cuando se aplica a muchos grupos, se ha propuesto el método de alineamiento como aproximación alternativa al estudio de la invarianza de medida. Sin embargo, existen interrogantes sobre su funcionamiento y la calidad de sus resultados. El objetivo de este trabajo ha sido estudiar y comparar el sesgo de las estimaciones y la potencia de ambos métodos. Para ello, se ha llevado a cabo un estudio de simulación Monte Carlo con 240 condiciones en las que se ha manipulado el tamaño muestral, el número de grupo, el tipo y la magnitud de no-invarianza. Los resultados han mostrado que el tamaño muestral tiene un claro efecto en ambos métodos sobre el sesgo y la potencia, mientras que el número de grupos, el tipo y la magnitud de no-invarianza parecen tener una influencia distinta en función del método y el tipo de parámetro. Teniendo estos resultados en cuenta, parece que cada uno de los métodos puede ser más adecuado en diferentes situaciones multigrupo, en función de los objetivos de la investigación que se quiera llevar a cabo.

Palabras clave: Invarianza de medida, análisis factorial multigrupo, método de alineamiento.

Abstract

Given the problems with the application of the Multigroup Factor Analysis to many groups, the alignment method has been proposed as an alternative approach to measurement invariance testing. However, there are certain questions regarding its performance and quality. The aim of this study was to examine and compare the bias of estimates and the power of both methods. In order to do so, a Monte Carlo simulation study was carried out, with 240 conditions where the sample size, the number of groups, the type and the magnitude of non-invariance were manipulated. The results have shown that the sample size has a clear effect in both methods in terms of bias and power, whereas the number of groups, the type and the magnitude of non-invariance have a different influence depending on the method and the type of parameter. Considering these results, it seems that each method could be more appropriate in distinct multigroup situations, according to the purpose of the research.

Keywords: Measurement invariance, multigroup factor analysis, alignment method.

Introducción

¿Qué es la invarianza de medida y por qué es importante?

Un aspecto importante que se debe tener en cuenta a la hora de evaluar las propiedades psicométricas de los test es la invarianza de medida, una propiedad que deben cumplir los instrumentos de evaluación para su aplicación legítima a distintos grupos o situaciones de medición. La invarianza implica que la respuesta a los ítems depende del nivel en el rasgo de las personas evaluadas independientemente del grupo al que pertenezcan (Mellenbergh, 1989). Cuando no hay invarianza de medida, la medida del rasgo no es válida en distintos grupos o situaciones, ya que está contaminada por varianza irrelevante debido a la influencia de factores que no son el objetivo de la medición.

La invarianza de medida es una propiedad fundamental y debe ser contrastada cuando las puntuaciones de los test se emplean para realizar comparaciones, seleccionar a personas o tomar decisiones (Borsboom, 2006; Chen, 2008; Meredith y Teresi, 2006), ya que permite asegurar que las puntuaciones observadas en la prueba y las conclusiones que se derivan de estas se deben únicamente al atributo de interés y no al efecto de otras características que no son el foco de la evaluación. Por ejemplo, en el caso de las comparaciones grupales de medias, que suele ser un objetivo habitual en Ciencias Sociales y de la Salud (Steinmetz, 2013), es muy importante que la medida sea invariante para que las diferencias en las medias de las puntuaciones observadas reflejen las diferencias reales en el rasgo medido y no se confundan con diferencias en características irrelevantes; o para que dichas diferencias no sean pasadas por alto (Borsboom, 2006; Chen, 2008; Meredith y Teresi, 2006).

A pesar de su importancia, es habitual asumir implícitamente que los instrumentos de medida utilizados son invariantes en los grupos que se desean evaluar y, en consecuencia, la invarianza de medida ni siquiera se pone a prueba (Borsboom, 2006; Byrne, 2016; van De Schoot, Schmidt, De Beuckelaer, Lek y Zondervan-Zwijenburg, 2015). Pese a su relevancia, esta asunción normalmente no se cumple, especialmente en el ámbito de las investigaciones transculturales y transnacionales, donde el objetivo generalmente es la comparación de diversos países o culturas y los test suelen funcionar de manera distinta en los diferentes grupos debido a una variedad de sesgos (Byrne, 2016; Davidov, Meuleman, Cieciuch, Schmidt y Billiet, 2014). Por todo ello, y dadas las consecuencias de su incumplimiento, es esencial comprobar que los instrumentos de

medida utilizados en la comparación o la toma de decisiones son invariantes en los grupos que son objeto de evaluación.

Análisis Factorial Confirmatorio Multigrupo (AFC-MG)

Dentro del marco de los modelos de variable latente y los Modelos de Ecuaciones Estructurales (SEM), el Análisis Factorial Confirmatorio Multigrupo (AFC-MG) es la técnica más conocida y utilizada tradicionalmente para poner a prueba la invarianza de medida (Millsap y Meredith, 2007; van De Schoot *et al.*, 2015). A nivel formal, el modelo se define del siguiente modo (Sörbom, 1974):

$$\mathbf{x}^{(g)} = \boldsymbol{\tau}_x^{(g)} + \boldsymbol{\Lambda}_x^{(g)} \boldsymbol{\xi}^{(g)} + \boldsymbol{\delta}^{(g)} \quad (g = 1, 2, \dots, G), \quad (1)$$

donde $\mathbf{x}^{(g)}$ es el vector de p variables observables en el grupo g , $\boldsymbol{\xi}^{(g)}$ es el vector de q factores latentes comunes en el grupo g que subyacen a las variables observables y que se distribuyen normalmente, $\boldsymbol{\Lambda}_x^{(g)}$ es la matriz de $p \times q$ pesos factoriales específicos de cada grupo que representan los pesos de regresión de cada variable observable en los factores comunes, $\boldsymbol{\tau}_x^{(g)}$ es el vector de p intersecciones específicas en g que representan el origen de la recta de regresión de cada indicador observable, y $\boldsymbol{\delta}^{(g)}$ es el vector de p errores de medida que siguen una distribución normal multivariada. Las medias y las varianzas-covarianzas de los indicadores observables se pueden obtener a partir de las medias y varianzas-covarianzas de los factores comunes de la siguiente forma:

$$\boldsymbol{\mu}_x^{(g)} = \boldsymbol{\tau}_x^{(g)} + \boldsymbol{\Lambda}_x^{(g)} \boldsymbol{\kappa}^{(g)}, \quad (2)$$

$$\boldsymbol{\Sigma}_{xx}^{(g)} = \boldsymbol{\Lambda}_x^{(g)} \boldsymbol{\Phi}^{(g)} \boldsymbol{\Lambda}_x^{(g)'} + \boldsymbol{\Theta}_\delta^{(g)}, \quad (3)$$

donde $\boldsymbol{\mu}_x^{(g)}$ y $\boldsymbol{\Sigma}_{xx}^{(g)}$ son el vector de p medias y la matriz $p \times p$ de varianzas-covarianzas de los ítems en el grupo g , respectivamente, $\boldsymbol{\kappa}^{(g)}$ es el vector de q medias factoriales, $\boldsymbol{\Phi}^{(g)}$ es la matriz $q \times q$ de varianzas-covarianzas factoriales, y $\boldsymbol{\Theta}_\delta^{(g)}$ es la matriz $p \times p$ de varianzas residuales, que se asumen independientes.

Teniendo en cuenta el modelo formulado en la Ecuación 1, la invarianza de medida implicaría que los parámetros que relacionan los indicadores observables con los factores latentes son iguales entre los distintos grupos. La puesta a prueba de la invarianza mediante AFC-MG conlleva un procedimiento jerárquico en el que se compara el ajuste de una serie de modelos anidados (Steenkamp y Baumgartner, 1998). Dichos modelos

son los que definen los distintos niveles de invarianza e imponen progresivamente más restricciones de igualdad en los parámetros analizados entre los grupos.

El primer nivel es la invarianza configural (Horn y McArdle, 1992), que sería el nivel más básico y el menos restrictivo. La invarianza configural implica que la estructura factorial (esto es, el número de factores y los ítems que constituyen cada factor) es la misma en todos los grupos; es decir, el patrón de pesos factoriales que son iguales a cero y los que son significativamente distintos de cero es igual en los grupos comparados (Horn y McArdle, 1992; Steenkamp y Baumgartner, 1998). Sin embargo, los parámetros del modelo se estiman libremente, salvo las restricciones mínimas necesarias para la identificación del modelo, y pueden ser distintos entre los grupos. Aunque la invarianza configural por sí sola no asegura que las comparaciones de las puntuaciones entre los distintos grupos sean válidas (Horn y McArdle, 1992), es el modelo sobre el cual se apoyan el resto de los niveles de invarianza.

El segundo nivel sería la invarianza métrica o invarianza débil (Horn y McArdle, 1992; Meredith, 1993; Steenkamp y Baumgartner, 1998). Se dice que hay invarianza métrica cuando todos los grupos, además de tener la misma estructura, tienen pesos factoriales iguales ($\Lambda_x^1 = \Lambda_x^2 = \dots = \Lambda_x^G = \Lambda_x$). Si se cumple esta igualdad, la unidad de medida es la misma en los grupos comparados; es decir, cambios de una unidad en el factor latente producen el mismo cambio en los ítems en todos los grupos. Por tanto, la invarianza métrica es necesaria cuando se quieren realizar comparaciones grupales de las varianzas factoriales y de las relaciones (covarianzas o coeficientes de regresión) con otras variables (Meredith y Teresi, 2006; Steenkamp y Baumgartner, 1998). De hecho, se ha mostrado que la falta de invarianza métrica hace que las estimaciones de los coeficientes de regresión pierdan precisión, tanto más cuanto mayor sea la cantidad de pesos distintos entre grupos (Chen, 2008).

El tercer nivel es la invarianza escalar o invarianza fuerte (Meredith, 1993; Steenkamp y Baumgartner, 1998), y suele ser el foco de los estudios en los que se realizan comparaciones entre grupos. Este nivel de invarianza implica que, además de ser iguales la estructura y los pesos factoriales, las intersecciones de los ítems son las mismas en todos los grupos ($\tau_x^1 = \tau_x^2 = \dots = \tau_x^G = \tau_x$). Cuando se cumple la invarianza escalar, tanto la métrica como el origen son los mismos en todos los grupos. Así pues, es necesaria cuando el objetivo es comparar las medias de los grupos, ya que asegura que las

diferencias en las medias de los ítems se deben a las diferencias en los factores comunes (Meredith y Teresi, 2006; Steenkamp y Baumgartner, 1998). Al igual que ocurría en la invarianza métrica, se ha observado que la falta de invarianza escalar sesga las estimaciones de las medias factoriales, tanto más cuantos más pesos e intersecciones distintas haya entre los grupos (Chen, 2008)

Por último, se encuentra la invarianza estricta (Meredith, 1993), que sería el nivel más restrictivo. Además de las restricciones de igualdad de los modelos anteriores, la invarianza estricta establece que las varianzas residuales son iguales en todos los grupos ($\Theta_{\delta}^1 = \Theta_{\delta}^2 = \dots = \Theta_{\delta}^G = \Theta_{\delta}$). Este nivel de invarianza implicaría que la fiabilidad de los ítems es la misma en los grupos. La invarianza estricta no se suele poner a prueba, ya que no es necesaria para comparar a los grupos en las medias o las varianzas-covarianzas factoriales (Steenkamp y Baumgartner, 1998).

El proceso de puesta a prueba de invarianza de medida mediante el AFC-MG es el siguiente. En primer lugar, se pone a prueba la invarianza configural, ajustando el mismo modelo factorial a todos los grupos a la vez, sin imponer restricciones de igualdad en los parámetros. Si el modelo de invarianza configural muestra buen ajuste a los datos, el siguiente paso es poner a prueba la invarianza métrica, imponiendo igualdad en todos los pesos factoriales, y se compara su ajuste frente al del modelo anterior. Para ello, se puede utilizar la diferencia de ajuste en distintos índices: chi-cuadrado (χ^2), Índice de Ajuste Comparativo (CFI), Raíz del Error Cuadrático Medio de Aproximación (RMSEA), etc. (Rutkowski y Svetina, 2014). Si la diferencia de ajuste entre los modelos de invarianza métrica y configural no resulta significativa, se puede decir que la invarianza métrica es compatible con los datos. El siguiente paso es poner a prueba la invarianza escalar, imponiendo la restricción adicional de igualdad en las intersecciones y comparando el ajuste de dicho modelo frente al de invarianza métrica, de manera similar al paso anterior. De nuevo, si la diferencia de ajuste no resulta significativa, se satisface la invarianza escalar en los grupos comparados (Steenkamp y Baumgartner, 1998).

Problemas del AFC-MG. Invarianza parcial

En determinadas circunstancias, como la comparación de un gran número de grupos, el AFC-MG plantea ciertos problemas. En este tipo de situaciones, la asunción de que la escala tiene la propiedad de invarianza es poco realista (Marsh *et al.*, 2018).

En primer lugar, la suposición de que todos los grupos comparados pertenecen a la misma población no suele cumplirse, especialmente en estudios transculturales o transnacionales a gran escala. Así pues, puede haber problemas incluso para llegar a un modelo configural adecuado para todos los grupos, ya que los modelos base para cada uno de ellos pueden ser distintos y la combinación de dichos modelos puede llevar a un modelo configural inadecuado (Byrne y van de Vijver, 2010).

En segundo lugar, el modelo que plantea el AFC-MG es muy restrictivo, imponiendo igualdad exacta en un gran número de parámetros, lo que frecuentemente suele llevar al rechazo de la hipótesis de invarianza, sobre todo cuando hay muchos grupos (Lek *et al.*, 2019). De hecho, la invarianza de medida, especialmente la invarianza escalar, no suele verse apoyada en estudios a gran escala (Davidov *et al.*, 2014; Davidov, Muthén y Schmidt, 2018; Desa, van de Vijver, Carstens y Schulz, 2019; Marsh *et al.*, 2018; van De Schoot *et al.*, 2015). Además, se ha observado que el rendimiento de los criterios de comparación de ajuste empeora a mayor número de grupos: por un lado, las diferencias en chi-cuadrado se vuelven más grandes y, por tanto, más significativas; por otro lado, parece ser que las diferencias en CFI y RMSEA requieren criterios más liberales cuando se comparan muchos grupos, especialmente para la invarianza métrica (Rutkowski y Svetina, 2014). Todo esto plantea un problema: no está claro si la falta de invarianza se debe a diferencias pequeñas pero insustanciales que son detectadas por el procedimiento, si realmente se debe a diferencias sustantivas en los modelos factoriales que apuntan a una verdadera falta de invarianza, o a una combinación de ambas circunstancias (Desa *et al.*, 2019; Byrne y van de Vijver, 2010). En cualquiera de estos casos, el rechazo de la invarianza de medida llevaría a la conclusión de que no puede hacerse una comparación válida entre los grupos, ya que, además, el AFC-MG no es capaz de lidiar con esta falta de equivalencia (Byrne y van de Vijver, 2010).

Una alternativa para lidiar con esa falta de invarianza detectada mediante el AFC-MG y poder realizar comparaciones grupales es la estrategia de la invarianza parcial (Byrne, Shavelson, y Muthén, 1989). La idea que subyace a la invarianza parcial es que se pueden realizar comparaciones grupales válidas cuando no hay invarianza de medida total, siempre y cuando al menos dos ítems sean totalmente invariantes (Byrne *et al.*, 1989; Steenkamp y Baumgartner, 1998). El procedimiento es similar al del AFC-MG, pero se apoya en los índices de modificación (MI) y los cambios esperados en los parámetros (EPC) para liberar las restricciones de igualdad, una a una, en aquellos

parámetros que más perjudican el ajuste del modelo (Byrne *et al.*, 1989; Steenkamp y Baumgartner, 1998).

A pesar de que la invarianza parcial se aplica con cierta frecuencia (Byrne y van de Vijver, 2010), es una estrategia que ha sido criticada. En primer lugar, desde un punto de vista práctico, parece que la invarianza parcial no es requisito suficiente para que las comparaciones sean válidas. Por ejemplo, Millsap y Kwok (2004) comprobaron que la falta de invarianza métrica y escalar total puede afectar de manera diferente en cada uno de los grupos a la precisión de la escala para seleccionar individuos, sobre todo, en términos de la sensibilidad del instrumento. Por otro lado, Steinmetz (2013) comprobó que en condiciones de invarianza escalar parcial las diferencias de medias observadas entre los grupos están sesgadas, tanto más cuantas más intersecciones distintas haya y menos ítems tenga la escala, y también se ve afectada la calidad del contraste de estas diferencias de medias.

Desde un punto de vista más técnico, es un procedimiento laborioso y poco práctico que puede requerir muchas modificaciones manuales del modelo, ya que el número de diferencias aumenta a más grupos, por lo que puede llegar a ser inviable con muchos grupos (Asparouhov y Muthén, 2014; Marsh *et al.*, 2018). Además, hay riesgo de que se produzca capitalización del azar (MacCallum, Roznowski y Necowitz, 1992) si las modificaciones se hacen en base a los datos en lugar de basándose en razones sustantivas (Byrne *et al.*, 1989; Steenkamp y Baumgartner, 1998). Esto podría dar lugar a un modelo con buen ajuste a los datos concretos, pero poco generalizable e incluso alejado del verdadero modelo (Asparouhov y Muthén, 2014; Marsh *et al.*, 2018).

Alternativa: invarianza aproximada. El alineamiento

En vista de que el AFC-MG y la invarianza parcial no parecen las estrategias más adecuadas para evaluar la invarianza de medida con muchos grupos, se han propuesto aproximaciones alternativas al estudio de la invarianza en estos contextos, entre las que se encuentra la invarianza aproximada. La idea subyacente es que las restricciones impuestas por los métodos exactos de invarianza, como el AFC-MG, son muy estrictas, ya que la hipótesis de partida es que las diferencias entre los parámetros de los grupos tienen que ser exactamente iguales a cero. Por ello, la invarianza aproximada impone restricciones menos estrictas, de manera que se parte de la hipótesis de que las diferencias son próximas a cero; es decir, se asume que las diferencias entre los parámetros son

pequeñas pero irrelevantes desde el punto de vista sustantivo para poder realizar comparaciones entre grupos (Lek *et al.*, 2019; van De Schoot *et al.*, 2015). Así pues, la invarianza aproximada representa una alternativa más realista a las situaciones que requieren la comparación de muchos grupos.

Dentro del marco de la invarianza aproximada, Asparouhov y Muthén (2014) desarrollaron recientemente el método de alineamiento (*alignment method*), un procedimiento que automatiza y simplifica la puesta a prueba de la invarianza de medida. Este método estima los parámetros del modelo (pesos, intersecciones, medias y varianzas factoriales) para cada grupo, permitiendo la comparación de medias y varianzas factoriales entre ellos a la vez que optimiza la invarianza de los pesos y las intersecciones.

Para ello, el procedimiento parte del modelo del AFC-MG:

$$y_{ijg} = \nu_{jg} + \lambda_{jg}\eta_{ig} + \varepsilon_{ijg}, \quad (4)$$

donde $j = 1, \dots, J$ se refiere al ítem, $g = 1, \dots, G$ indica el grupo e $i = 1, \dots, N_g$ se refiere al individuo concreto en el grupo g . La variable latente η_{ig} se asume que se distribuye según $N(\alpha_g, \psi_g)$ y los errores de medida ε_{ijg} se distribuyen según $N(0, \theta_{pg})$ (Asparouhov y Muthén, 2014).

En primer lugar, el método de alineamiento estima el modelo configural, donde los pesos factoriales y las intersecciones son estimados libremente y sin restricciones de igualdad, mientras que se fijan las medias factoriales a cero ($\alpha_g = 0$) y las varianzas factoriales a uno ($\psi_g = 1$) para todos los grupos. El modelo final alineado que se obtiene con este método tiene el mismo ajuste que el modelo configural, y sus parámetros se pueden expresar en términos de las medias y las varianzas factoriales, así como de los pesos y las intersecciones del modelo configural. Esto es así porque, al establecer que la media y la varianza del factor en el modelo configural son cero y uno, respectivamente, la media y la varianza de los ítems, según lo planteado en las Ecuaciones 2 y 3, se pueden obtener de la siguiente manera:

$$E(y_{jg}) = \nu_{jg} + \lambda_{jg}\alpha_g = \nu_{jg,0}, \quad (5)$$

$$V(y_{jg}) = \lambda_{jg}^2\psi_g = \lambda_{jg,0}^2 \quad (6)$$

Así pues, las estimaciones de los pesos ($\lambda_{jg,0}$) y las intersecciones ($\nu_{jg,0}$) del modelo configural pueden expresarse del siguiente modo:

$$\lambda_{jg,0} = \lambda_{jg} \sqrt{\psi_g}, \quad (7)$$

$$v_{jg,0} = v_{jg} + \frac{\lambda_{jg,0}}{\sqrt{\psi_g}} \alpha_g, \quad (8)$$

con lo que los parámetros del modelo final alineado se pueden obtener de la siguiente manera:

$$\lambda_{jg,1} = \frac{\lambda_{jg,0}}{\sqrt{\psi_g}}, \quad (9)$$

$$v_{jg,1} = v_{jg,0} - \frac{\lambda_{jg,0}}{\sqrt{\psi_g}} \alpha_g \quad (10)$$

Según las Ecuaciones 9 y 10, Asparouhov y Muthén (2014) plantean que a partir de cada par de medias y varianzas factoriales se obtienen distintos valores para los pesos y las intersecciones alineadas que dan lugar a la misma verosimilitud que el modelo configural. Por tanto, lo que interesa es escoger aquellas medias y varianzas que maximicen la invarianza en los pesos y las intersecciones. De manera formal, el objetivo es minimizar la función de simplicidad/pérdida total, \mathbf{F} , respecto a las medias y las varianzas. La función \mathbf{F} acumula la diferencia entre cada par de pesos e intersecciones para cada par de grupos:

$$\mathbf{F} = \sum_j \sum_{g_1 < g_2} w_{g_1, g_2} f(\lambda_{jg_1,1} - \lambda_{jg_2,1}) + \sum_p \sum_{g_1 < g_2} w_{g_1, g_2} f(v_{jg_1,1} - v_{jg_2,1}) \quad (11)$$

La diferencia entre cada par de estimaciones está influida por el peso del tamaño muestral, w_{g_1, g_2} , y está escalada por la función de componente de pérdida (*component loss function*, CLF) f ,

$$f(x) = \sqrt{\sqrt{x^2 + 0.01}}, \quad (12)$$

cuya lógica es la misma que la de la rotación en el Análisis Factorial Exploratorio (AFE). Así pues, la función \mathbf{F} se minimiza cuando la mayoría de las diferencias entre los parámetros son pequeñas y solo unas pocas son grandes (es decir, muchos parámetros son aproximadamente invariantes y solo unos pocos presentan gran no-invarianza), en lugar de con un gran número de diferencias de tamaño medio (Asparouhov y Muthén, 2014).

Hay dos tipos de alineamiento, en función de las restricciones realizadas para la identificación del primer grupo. En primer lugar, se puede restringir únicamente ψ_1 (mediante $\psi_1 \times \dots \times \psi_g = 1$), lo que se conoce como alineamiento libre (*FREE*

alignment), en el que se estiman $2G - 1$ medias y varianzas factoriales. Por otro lado, también se puede restringir $\alpha_1 = 0$, dando lugar al alineamiento fijo (*FIXED alignment*). Con ambos métodos, una vez que se han estimado las medias y las varianzas factoriales, se calculan los pesos y las intersecciones del modelo alineado, a partir de las Ecuaciones 9 y 10. Además, la estimación de parámetros se puede realizar con Máxima Verosimilitud (ML) o mediante estimación bayesiana (Asparouhov y Muthén, 2014).

Una vez estimados todos los parámetros del modelo alineado, se lleva un procedimiento *ad-hoc* basado en la comparación por pares que proporciona información sobre qué grupos resultan no-invariantes en cada uno de los pesos y de las intersecciones. Este procedimiento primero determina para cada peso e intersección el conjunto inicial invariante. Para ello, se comparan los valores del parámetro entre cada par de grupos, de manera que, si la diferencia no resulta significativa ($p > .01$), dichos grupos se conectan. Estas comparaciones se repiten hasta que se llega al mayor conjunto posible de grupos conectados para dicho parámetro, es decir, el conjunto inicial invariante. Posteriormente, se calcula el parámetro (peso o intersección) medio del conjunto inicial invariante y se compara con el valor de dicho parámetro en cada uno de los grupos. Si la diferencia no resulta significativa ($p > .001$), el grupo pasa a formar parte del conjunto invariante, y si resulta significativa se elimina de este. El proceso se repite de forma iterativa hasta que no se añaden ni se eliminan más grupos del conjunto invariante (Asparouhov y Muthén, 2014). Aquellos grupos que no hayan sido añadidos al conjunto invariante final serían no-invariantes para dicho parámetro, mientras que los grupos que forman parte de este conjunto se considerarían invariantes.

Aplicaciones del alineamiento

Algunos autores han aplicado el método de alineamiento a datos reales y han defendido su utilidad práctica cuando se comparan muchos grupos, frente al AFC-MG y los problemas que plantea (Byrne y van de Vijver, 2017; Muthén y Asparouhov, 2013), y en el contexto de la Teoría de Respuesta al Ítem (TRI) cuando se aplica a ítems dicotómicos (Muthén y Asparouhov, 2014). Además, desde que se propuso esta técnica hasta hoy, y especialmente a partir de 2017, diversos autores en numerosos campos de investigación de las Ciencias Sociales y de la Salud han aplicado el método de alineamiento en sus estudios.

Se realizó una revisión de dichos estudios, con el propósito de examinar en qué condiciones se ha empleado el método de alineamiento en la práctica. Por tanto, se llevó a cabo una búsqueda en la base de datos PsycINFO de los trabajos publicado entre 2014 y marzo de 2019, empleando las siguientes palabras clave: “*alignment method*”, “*alignment optimization*”, “*approximate measurement invariance*”, “*measurement invariance*”, “*multigroup factor analysis*” y “*cross-cultural differences*”. Con el fin de incluir todas las investigaciones posibles, se realizó también una búsqueda en Google Scholar de los estudios que citaron el trabajo original de Asparouhov y Muthén (2014), cuyas palabras clave no coincidían con las de la búsqueda anterior pero que emplearon el alineamiento. El criterio de inclusión en la revisión fue únicamente que el estudio en cuestión empleara el método de alineamiento bien para evaluar la invarianza de medida, bien para comparar las medias entre grupos, o ambos. Finalmente, se revisaron 50 trabajos que cumplieran este criterio de inclusión. Por limitaciones de espacio, únicamente se comentan de forma breve una serie de conclusiones generales de las características y los resultados de los trabajos, citando algunos estudios destacados. No obstante, en el Apéndice A del trabajo se adjunta la Tabla A1 con algunas características de interés de todos los trabajos que han aplicado el método de alineamiento, incluyendo el número de grupos bajo estudio, la muestra total o submuestras grupales, la medida empleada (con el número de ítems y de factores) y las técnicas de invarianza y de comparación de medias empleadas, así como el tipo de alineamiento y el método de estimación utilizados (si alguna de estas características no se aportaba de forma explícita en el trabajo, la información no ha sido incluida en la Tabla A1).

A partir de esta revisión, se ha observado que el método de alineamiento se ha empleado en la práctica bien con el objetivo principal de evaluar la invarianza de medida de algún instrumento o comparar las puntuaciones entre grupos, bien como paso previo a otros objetivos (estudio de relaciones entre variables, desarrollo de escalas, etc.) para comprobar que las medidas utilizadas son equivalentes entre grupos. El número de grupos en los que se ha aplicado la técnica es amplio, abarcando desde 2 grupos (por ejemplo, de Bondt y Van Petegem, 2015) hasta 92 grupos en total (Munck, Barber y Torney-Purta, 2018). Los tamaños muestrales grupales en que se ha empleado también han sido muy diversos, abarcando desde muestras inferiores a 100 participantes (por ejemplo, Evers *et al.*, 2017) hasta muestras con más de 4000 evaluados (por ejemplo, Barber y Ross, 2018).

Por otro lado, se ha empleado el alineamiento para evaluar medidas de diversa complejidad. Lo más habitual es que las escalas incluyeran únicamente un factor latente (por ejemplo, Jang *et al.*, 2017; Lomazzi, 2018), pero se ha llegado a emplear hasta con instrumentos de 9 factores (Elsworth, Beauchamp y Osborne, 2016). El número de ítems por factor suele ser, generalmente, entre 3 y 6, aunque se han llegado a examinar escalas con 10 ítems o más por factor (por ejemplo, McGrath, 2015). Dichos ítems han sido, por norma general, politómicos, siendo lo más habitual que el número de categorías de respuesta fuera en torno a cinco.

Además, aunque el alineamiento se podía emplear como única técnica, normalmente se ha utilizado junto con otras técnicas de evaluación de invarianza, sobre todo, el AFC-MG. Respecto a los resultados de los estudios, en general parece haber bastante consistencia sobre que la invarianza métrica es más habitual que la invarianza escalar, ya que se suele encontrar que los pesos resultan más invariantes entre los distintos grupos bajo estudio que las intersecciones. Estos resultados van en la línea de las críticas sobre el AFC-MG comentadas anteriormente que evidencian la dificultad para encontrar instrumentos que tengan propiedades invarianza muy restrictivas, especialmente escalar, cuando se compara un gran número de grupos.

Estudios de simulación. Ventajas e inconvenientes del alineamiento

Dado que el alineamiento parece ser un procedimiento con un gran potencial aplicado, es necesario conocer si los resultados que proporciona son adecuados. Varios autores han comprobado mediante estudios de simulación el funcionamiento del método de alineamiento y la calidad de sus resultados. No obstante, dichos estudios, que se comentan a continuación, son hasta el momento escasos, dado que la técnica es muy reciente.

Asparouhov y Muthén (2014), además de proponer el método de alineamiento, realizaron un estudio de simulación y comprobaron que las estimaciones de los parámetros eran bastante precisas en casi todas las condiciones. No obstante, el sesgo aumentaba a mayor cantidad de no-invarianza (es decir, número de parámetros que presentan no-invarianza) y menor tamaño muestral. Además, cuando se comparaban muchos grupos, la precisión de las estimaciones se veía más afectada cuanto más pequeña era la muestra. Asimismo, el alineamiento libre parecía dar mejores estimaciones que el alineamiento fijo salvo cuando solo había dos grupos o los parámetros eran totalmente

invariantes. Además, observaron que la estimación bayesiana parece ser mejor que la estimación con ML. En otro estudio de simulación, los autores comprobaron que la correlación entre las medias simuladas y las medias estimadas por el alineamiento era muy alta, por lo que parece que el método estima de manera bastante precisa las medias de los grupos (Muthén y Asparouhov, 2013).

Por otro lado, Marsh *et al.* (2018) estudiaron la precisión de las estimaciones de las medias con el alineamiento en comparación con las del AFC-MG tanto con invarianza total como parcial. Los autores comprobaron que el alineamiento daba lugar a estimaciones más precisas que la estrategia de invarianza total y, sobre todo, que la de invarianza parcial en todas las condiciones estudiadas. Además, el sesgo de las estimaciones fue menor a mayor cantidad de invarianza y a mayor tamaño muestral para las tres técnicas.

Por su parte, Kim, Cao, Wang y Nguyen (2017) investigaron la eficacia de distintos procedimientos para detectar no-invarianza, entre ellos, el alineamiento. Estos autores comprobaron que, en términos generales, el método mostraba un buen rendimiento, especialmente el alineamiento fijo. No obstante, parecía detectar mejor los grupos invariantes que los no-invariantes, sobre todo con tamaño muestral más pequeño. Además, parecía funcionar mejor cuando existía invarianza aproximada, con diferencias pequeñas entre los parámetros grupales, que en condiciones de magnitud de no-invarianza (es decir, diferencia en el valor de los parámetros) grande.

Pokropek, Davidov y Schmidt (2019) evaluaron la calidad de las estimaciones de las medias latentes y los coeficientes de regresión obtenidas mediante diferentes métodos, incluido el alineamiento, bajo condiciones de invarianza parcial, invarianza aproximada e invarianza aproximada parcial. Estos autores comprobaron que, en condiciones de invarianza parcial, el método de alineamiento funcionaba adecuadamente, incluso algo mejor que el AFC-MG, cuando el número de ítems y grupos no-invariantes era limitado, pero las estimaciones eran menos precisas a mayor cantidad de no-invarianza, funcionando peor que la estrategia de invarianza parcial. En condiciones de invarianza aproximada, todos los métodos funcionaban de manera similar. Por su parte, en condiciones de invarianza aproximada parcial, el alineamiento funcionó de manera bastante similar al AFC-MG, y solo proporcionaba estimaciones adecuadas en un número muy limitado de condiciones, empeorando su funcionamiento a mayor número de grupos e ítems no-invariantes.

Por otro lado, Flake y McCoach (2018) estudiaron el funcionamiento y la calidad del método de alineamiento con ítems politómicos. En general, comprobaron que los resultados respecto a la precisión de las estimaciones de los parámetros eran buenos en la mayoría de las condiciones. No obstante, se observó una tendencia de aumento del sesgo de las estimaciones a mayor cantidad y magnitud de no-invarianza (y a más grupos, en el caso de las medias factoriales). Además, el tipo de no-invarianza afectaba a la localización de dicha tendencia, de manera que si había no-invarianza en los pesos, se veía afectada la precisión en los pesos y las varianzas, mientras que si los umbrales no eran invariantes, se incrementaba el sesgo en los umbrales y las medias. Respecto al procedimiento *ad-hoc*, comprobaron que, si bien los grupos invariantes eran bien detectados, resultaba demasiado conservador a la hora de detectar los grupos no-invariantes, incluso con magnitud de no-invarianza grande, especialmente para los pesos.

Dentro del marco de la TRI, Finch (2016) comparó el funcionamiento de diversas técnicas para detectar Funcionamiento Diferencial del Ítem (DIF) uniforme en ítems dicotómicos, incluyendo el método de alineamiento. Los resultados mostraron que, en general, el alineamiento funcionaba mejor que el resto de las técnicas, con una mejor combinación de tasa de error tipo I y potencia. No obstante, ambos resultados empeoraban a menor tamaño muestral, y la potencia resultó mayor a más nivel de DIF. La ratio grupal y el número de grupos también parecían influir en la detección de DIF. Por otro lado, las estimaciones de los parámetros de dificultad del ítem con DIF resultaron más sesgadas con el método de alineamiento cuando existía diferencia de medias latentes entre los grupos y mayor nivel de DIF.

Así pues, teniendo en cuenta los resultados de todos estos estudios, parece que el método de alineamiento es un procedimiento prometedor que, además de automatizar y simplificar el estudio de invarianza con un gran número de grupos, puede funcionar igual o incluso mejor que otras técnicas, como el AFC-MG, bajo diferentes condiciones. En general, parece que el alineamiento estima de forma bastante precisa los parámetros del modelo, aunque el sesgo y la eficiencia empeoran cuanto mayor es la magnitud o la cantidad de no-invarianza. Asparouhov y Muthén (2014) plantearon que el método de alineamiento trabaja bajo la asunción implícita de que la mayor parte de las diferencias entre los parámetros de los grupos son pequeñas e irrelevantes, de manera que, si dicha asunción no se cumple, los resultados del procedimiento pueden ser inadecuados. Esta limitación ha sido señalada de manera consistente por otros autores: para que el

procedimiento funcione de forma adecuada, es necesario que la mayoría de los parámetros sean aproximadamente invariantes (Kim *et al.*, 2017; Muthén y Asparouhov, 2013; Pokropek *et al.*, 2019). No obstante, tampoco está claro cuánta no-invarianza es asumible. A modo de guía, se ha propuesto que hasta un 25% de no-invarianza en los parámetros los resultados del alineamiento son confiables (Muthén y Asparouhov, 2014), aunque dicha regla no tiene en cuenta la localización y la magnitud concreta de las diferencias entre los parámetros (Kim *et al.*, 2017).

Por otro lado, parece que las estimaciones también son menos precisas cuanto menor es el tamaño muestral y mayor el número de grupos, sobre todo si las muestras son pequeñas. Así pues, se ha propuesto que es un método adecuado salvo que el número de grupos sea muy grande (por ejemplo, mayor que 100) o las muestras sean muy pequeñas, donde otras técnicas, como los modelos multinivel, podrían dar mejores resultados (Muthén y Asparouhov, 2013).

Una de las grandes ventajas que presenta el método de alineamiento es que proporciona información de manera automática sobre qué grupos resultan no-invariantes en cada uno de los parámetros de los ítems mediante el procedimiento *ad-hoc*, lo que a su vez permite identificar qué tipo de parámetro y qué ítems son más o menos invariantes (Asparouhov y Muthén, 2014). Este procedimiento parece detectar de manera adecuada los grupos que son realmente invariantes, con un control de la tasa de error tipo I adecuado. Sin embargo, parece que su potencia a la hora de detectar los grupos no-invariantes es baja, resultando demasiado conservador. Así pues, a pesar de que el procedimiento *ad-hoc* es uno de los grandes atractivos del método de alineamiento, ha recibido algunas críticas, ya que sus resultados dependen no solo de la verdadera magnitud de las diferencias, sino del tamaño muestral y el nivel de significación (Flake y McCoach, 2018; Marsh *et al.*, 2018).

El presente trabajo. Objetivo e hipótesis

El método de alineamiento es una alternativa prometedora al clásico AFC-MG. Sin embargo, se han realizado pocos estudios comparativos sobre su funcionamiento y la calidad de sus resultados, por lo que todavía no está claro en qué medida es un procedimiento adecuado para el estudio de invarianza. Además, solo dos estudios (Marsh *et al.*, 2018; Pokropek *et al.*, 2019) han comparado directamente el método de

alineamiento con el AFC-MG, por lo que tampoco queda claro hasta qué punto el alineamiento resulta beneficioso frente a la aproximación clásica al estudio de invarianza.

Por tanto, la finalidad de este trabajo es investigar el funcionamiento y la calidad de las estimaciones del AFC-MG y del método de alineamiento bajo diferentes condiciones. Dado que la magnitud de no-invarianza en los parámetros, el número de grupos y el tamaño muestral parecen ser factores de gran relevancia, se decidió estudiar su influencia sobre los resultados de ambos métodos. Además, parece que la localización de la no-invarianza puede influir en los resultados y su efecto ha sido poco investigado. Así pues, los objetivos son: (a) comprobar el sesgo de las estimaciones del AFC-MG y del alineamiento en función del número de grupos, el tamaño muestral, la magnitud y la localización de la no-invarianza; y (b) estudiar la potencia de ambos métodos en función de dichas variables independientes. Las hipótesis de partida son: (1) el sesgo aumentará a mayor magnitud de no-invarianza, mayor número de grupos y menor tamaño muestral en ambos métodos; (2) el sesgo se verá afectado de forma diferente en función de la localización de la no-invarianza en ambos métodos; (3) el método de alineamiento proporcionará estimaciones más precisas que el AFC-MG; y (4) la potencia del AFC-MG aumentará a mayor magnitud de no-invarianza, mayor número de grupos y mayor tamaño muestral.

Dado que el procedimiento *ad-hoc* del método de alineamiento ha sido criticado, el trabajo tiene el objetivo adicional de proponer un procedimiento de contraste basado en tamaños del efecto en lugar de significación estadística y comprobar su rendimiento. Los tamaños del efecto tienen la ventaja de que no dependen tanto del tamaño de la muestra (Ferguson, 2009) y ponen el foco en la relevancia sustantiva de las diferencias (Borsboom, 2006), por lo que parecen una estrategia atractiva dentro del marco de la invarianza aproximada y, más concretamente, para el método de alineamiento. Así pues, se plantea la hipótesis de que (5) la potencia del alineamiento aumentará a mayor magnitud de no-invarianza, pero se mantendrá constante independientemente del tamaño muestral. Respecto al número de grupos, no se planteó ninguna hipótesis concreta, ya que los resultados de los estudios no parecen arrojar una tendencia clara de su influencia en la potencia.

Método

Para llevar a cabo los objetivos planteados, se realizó un estudio de simulación Monte Carlo. Este tipo de estudios es ideal para evaluar la calidad de las estimaciones y la potencia de los procedimientos estadísticos, ya que permiten extraer múltiples muestras a partir de unos parámetros poblacionales hipotetizados y aplicar el procedimiento de interés en cada una de esas muestras (Muthén, 2002). En las simulaciones, se manipularon los siguientes factores: el número de grupos (4 condiciones), el tamaño muestral de los grupos (3 condiciones), el tipo de no-invarianza (2 condiciones), la magnitud de no-invarianza (4 y 5 condiciones) y el método de análisis de invarianza (2 condiciones). Para este trabajo, se realizaron 1000 réplicas por cada una de las 240 condiciones de estudio resultantes. Se emplearon 1000 réplicas porque es un valor utilizado en otros estudios de simulación (por ejemplo, Finch, 2016) y, además, es superior a las 500 réplicas recomendadas por Muthén (2002) para obtener resultados estables.

El modelo factorial empleado, los valores poblacionales simulados y las distintas condiciones fueron escogidas de forma que fueran similares tanto a los estudios de simulación publicados como a los estudios prácticos revisados que han aplicado el método de alineamiento, con el fin de que los resultados obtenidos sean comparables y tengan relevancia a nivel práctico (si se desea más información sobre las características de dichos estudios en relación con las condiciones de simulación, se recomienda consultar la Tabla A1 de Apéndice A). La generación y el análisis de los datos se llevaron a cabo con el programa RStudio para Windows, empleando los paquetes lavaan (Rosseel, 2012), semTools (Jorgensen *et al.*, 2018) y sirt (Robitzsch, 2019).

Generación de datos

Los datos se generaron de acuerdo con un modelo tau-equivalente de 6 ítems continuos y un factor latente, bajo la asunción de normalidad. Se escogió un modelo de una variable latente porque los estudios de simulación han empleado generalmente un factor (por ejemplo, Asparouhov y Muthén, 2014; Finch, 2016; Kim *et al.*, 2017; Marsh *et al.*, 2018; Muthén y Asparouhov, 2013; Pokropek *et al.*, 2019). Además, en la gran mayoría de los estudios revisados que han aplicado el alineamiento han utilizado también medidas de un factor. Por otro lado, se decidió que fueran ítems continuos y normales para poder comparar los resultados con otros estudios de simulación que emplearon estas mismas características (por ejemplo, Asparouhov y Muthén, 2014; Kim *et al.*, 2017;

Marsh *et al.*, 2018; Muthén y Asparouhov, 2013). Se escogieron 6 indicadores porque se ha comprobado que un número escaso de ítems (por ejemplo, 3 y 4) puede no ser suficiente para que las estimaciones sean precisas y eficientes en el AFC-MG y el alineamiento, incluso en condiciones favorables (Pokropek *et al.*, 2019). Además, dado que la mayoría de los estudios aplicados ha evaluado medidas de pocos ítems por factor, generalmente entre 3 y 6, se consideró que utilizar 6 indicadores era una opción adecuada.

En cuanto a los valores de los parámetros, todos los pesos factoriales se fijaron a .8 ($\lambda_{jg} = .8$); las intersecciones de los ítems se fijaron a 0 ($v_{jg} = 0$) y todas las medias y varianzas factoriales de los grupos se fijaron a 0 y 1, respectivamente ($\alpha_g = 0$, $\psi_g = 1$). Dichos valores han sido utilizados en otros estudios de simulación que han evaluado el AFC-MG y el alineamiento (por ejemplo, Kim *et al.*, 2017; Steinmetz, 2013) y facilitan la identificación del modelo y la estandarización de los parámetros.

Número de grupos

El efecto de la cantidad de grupos se estudió variando el número de grupos estudiados entre 4, 8, 16 y 32. Estos valores permiten poner a prueba situaciones multigrupo que implican grados diferentes de complejidad y son bastante consistentes con el resto de estudios de simulación, ya que generalmente se han utilizado 20 grupos o menos. No obstante, lo más relevante es que estos números de grupos se corresponden con valores habituales en las investigaciones aplicadas. Aunque el alineamiento se ha aplicado a un rango amplio de grupos, la mayoría de los trabajos comparaban 35 grupos o menos, siendo los valores más habituales de 2 a 5 grupos y de 20 a 30 grupos.

Tamaño muestral

Se varió el tamaño de la muestra grupal entre 250, 500 y 1000. Dentro de cada condición, el tamaño muestral simulado fue el mismo en todos los grupos. Estas tres condiciones resultan adecuadas porque representan situaciones en las que se cuenta con una muestra pequeña, media y grande, respectivamente. Además, son valores que se han empleado en otros estudios de simulación y que representan muestras grupales habituales en las investigaciones aplicadas. Se decidió establecer el tamaño muestral en 250 y no en un valor menor porque se ha recomendado emplear el AFC y los modelos SEM con muestras superiores a los 200 participantes (Kline, 2016); y los estudios revisados han utilizado habitualmente muestras superiores a dicha recomendación.

Tipo y magnitud de no-invarianza

Se estudiaron dos tipos de no-invarianza: la no-invarianza en pesos y la no-invarianza en intersecciones, de manera similar a Flake y McCoach (2018). Así pues, en un conjunto de simulaciones, se estableció que las diferencias estuvieran únicamente en los pesos factoriales, siendo iguales las intersecciones; mientras que, en otro conjunto de simulaciones, las diferencias estaban localizadas en las intersecciones, siendo iguales los pesos factoriales. En las condiciones de no-invarianza en los pesos, se estableció que el peso del ítem 2 en el grupo 3 ($\lambda_{2,3}$) fuese no-invariante; mientras que en las condiciones de no-invarianza en las intersecciones, se simuló que fuera distinta la intersección del ítem 4 en el grupo 2 ($\nu_{4,2}$). Para generar las condiciones con 8, 16 y 32 grupos, se copió la estructura de la condición de 4 grupos tantas veces como fuera necesario, de forma similar a como se ha hecho en otros estudios de simulación (por ejemplo, Asparouhov y Muthén, 2014; Flake y McCoach, 2018). Así, por ejemplo, en la condición de 8 grupos con no-invarianza en los pesos, todos los pesos factoriales eran invariantes salvo los del ítem 2 en el grupo 3 ($\lambda_{2,3}$) y en el grupo 7 ($\lambda_{2,7}$) que, además, toman el mismo valor. Por tanto, se mantuvo constante el número de ítems no-invariantes en la escala a uno (lo que representa un 16.67% de no-invarianza), así como el porcentaje de grupos no-invariantes al 25%.

Respecto a la magnitud de las diferencias en los parámetros, dependiendo del tipo de no-invarianza, se estudiaron rangos de diferencias distintos. En el caso de la no-invarianza en los pesos, se simuló que las diferencias entre los pesos invariantes y no-invariantes fueran de .1 a .4, en intervalos de cambio de .1, con lo que los pesos no-invariantes podían valer .7; .6; .5 o .4. En el caso de la no-invarianza en las intersecciones, se simuló que las diferencias entre las intersecciones invariantes y las no-invariantes fueran de .1 a .5, también en intervalos de cambio de .1, con lo que las intersecciones no-invariantes podían tomar valores de .1; .2; .3; .4 o .5. Además, se incluyó la situación de invarianza completa, en la que las diferencias entre los parámetros eran exactamente iguales a cero. Estos rangos incluyen diferencias entre los parámetros que van desde diferencias nulas hasta magnitudes pequeñas, moderadas o grandes.

Método de evaluación de invarianza

La evaluación de la invarianza se llevó a cabo utilizando el AFC-MG y el método de alineamiento. En ambos métodos se empleó como método de estimación Máxima

Verosimilitud (ML), ya que se simularon ítems continuos. Además, los estudios de simulación sobre el alineamiento generalmente han empleado ML como método de estimación, así como la gran mayoría de las investigaciones aplicadas. Por ello, parece que tiene más interés aplicado estudiar la estimación ML del alineamiento que la bayesiana. Por el mismo motivo, se empleó el alineamiento fijo en lugar del libre. A pesar de que se ha apuntado a que el alineamiento libre podría ser mejor que el fijo, puede dar problemas de identificación. Si esto ocurre, es más recomendable usar el alineamiento fijo (Asparouhov y Muthén, 2014). De hecho, esta circunstancia surgió en un gran número de los trabajos aplicados revisados. Por tanto, se consideró más relevante comprobar el funcionamiento del alineamiento fijo que el del libre.

Medidas de la calidad de los resultados

Para evaluar el funcionamiento y la calidad de los resultados de ambos procedimientos en función de las distintas variables independientes, se estudiaron dos variables dependientes: por un lado, el sesgo de las estimaciones de los parámetros y, por otro lado, la potencia del método para detectar la falta de invarianza.

El sesgo de las estimaciones se refiere a la diferencia entre el parámetro estimado por el procedimiento y el parámetro simulado real a través de las réplicas. En este trabajo, se investiga el sesgo de los pesos (λ_{jg}), las intersecciones (ν_{jg}), las medias (α_g) y las varianzas factoriales (ψ_g) para cada uno de los grupos. Como medida de la magnitud del sesgo, se empleó la Raíz del Error Cuadrático Medio (RMSE):

$$RMSE = \sqrt{\frac{\sum(\hat{\theta} - \theta)^2}{Nr}}, \quad (13)$$

donde $\hat{\theta}$ es la estimación del parámetro en la réplica concreta, θ es el verdadero valor poblacional simulado del parámetro, y Nr es el número total de réplicas (en este caso, $Nr = 1000$). RMSE toma valores más pequeños cuanto menor es el sesgo; es decir, su valor es menor cuanto menor es la diferencia entre la estimación del parámetro y su valor real. Por tanto, cuanto más bajo sea RMSE, más precisas serán las estimaciones.

Para cada una de las condiciones de simulación, se calculó el RMSE de cada parámetro tal y como aparece en la Ecuación 13 y, a partir de estos resultados, se promedió el RMSE en cada condición para cada tipo de parámetro estudiado. De esta forma, por cada condición de simulación se obtuvieron cuatro valores promedio globales

de RMSE, uno por cada tipo de parámetro (λ_{jg} , ν_{jg} , α_g y ψ_g). Sin embargo, el promedio global puede no ser suficientemente informativo, ya que no distingue entre los ítems o los grupos invariantes y los no-invariantes y cabe la posibilidad de que algunos ítems o grupos estén mejor estimados que otros. Además, esta distinción no se ha hecho en otros estudios de simulación, por lo que parece relevante incluirla para poder comprobar si los métodos funcionan de manera diferente dependiendo de si los parámetros se refieren a elementos invariantes o no-invariantes. Por tanto, aparte de calcular el promedio global de RMSE, se calculó el RMSE promedio según el tipo de ítem y grupo. Para los pesos y las intersecciones se calculó, por un lado, el valor promedio de RMSE para los ítems invariantes y, por otro lado, el RMSE promedio para los ítems no-invariantes. Para las medias y las varianzas se calcularon los valores promedio de RMSE para los grupos invariantes y para los grupos no-invariantes. Así pues, por cada una de las condiciones de simulación, se obtuvo un total de 12 valores de RMSE, tres por cada tipo de parámetro: el promedio global de RMSE para dicho parámetro, el RMSE promedio para los ítems o grupos invariantes y el RMSE promedio para los ítems o grupos no-invariantes.

Por otro lado, la potencia se refiere a la proporción de réplicas en las que la hipótesis nula es rechazada (Muthén y Muthén, 1998-2017). Cuanto más cercana sea a uno, mayor es la capacidad de la técnica para detectar la falta de invarianza. Como para cada uno de los métodos de evaluación de invarianza se utilizó un procedimiento de contraste distinto, la potencia no tiene exactamente el mismo significado para el AFC-MG que para el alineamiento.

En el caso del AFC-MG, el contraste de hipótesis se realizó empleando la diferencia de chi-cuadrado en la comparación de modelo anidados, empleando un nivel de significación $\alpha = .05$. Así pues, la hipótesis nula de invarianza se rechazaba si la diferencia de chi-cuadrado resultaba significativa ($p < .05$), independientemente de si se estaba evaluando invarianza métrica o invarianza escalar. Por tanto, en este caso la potencia es la proporción de réplicas en las que la comparación de modelos resulta significativa.

En el caso del alineamiento, se empleó un contraste de hipótesis basado en tamaños del efecto. La estrategia seguida para este contraste es parecida al procedimiento *ad-hoc* propuesto por Asparuhov y Muthén (2014), en el sentido de que se compara el parámetro medio (peso o intersección) del conjunto invariante frente los parámetros no-

invariantes para el ítem con falta de invarianza. Sin embargo, dado que se conoce de antemano qué parámetros forman parte del conjunto invariante y cuáles no, no se llevó a cabo un procedimiento iterativo previo de comparaciones por pares para formar los conjuntos, sino que se calculó directamente la estimación media en cada uno de ellos. Además, salvo en las condiciones de 4 grupos, se decidió comparar con la media del conjunto no-invariante en lugar de con cada parámetro no-invariante por separado para simplificar el proceso, dado que todos los parámetros no-invariantes tomaban el mismo valor en dicha condición. Por ejemplo, en la condición de 8 grupos con no-invarianza en los pesos, se conoce de antemano que para el ítem 2, que es el ítem no-invariante, $\lambda_{2,1}$, $\lambda_{2,2}$, $\lambda_{2,4}$, $\lambda_{2,5}$, $\lambda_{2,6}$ y $\lambda_{2,8}$ son invariantes y valen .8, mientras que $\lambda_{2,3}$ y $\lambda_{2,7}$ son no-invariantes y su valor es el mismo (por ejemplo, .4). En este caso, los seis primeros parámetros formarían el conjunto invariante y los dos últimos formarían el conjunto no-invariante, y se calcularían directamente sus respectivas medias para realizar el contraste.

Una vez calculadas las medias, se utiliza un criterio basado en tamaños del efecto para determinar si se detecta o no la falta de invarianza en el parámetro: si la diferencia entre las medias de ambos conjuntos (o entre la media del conjunto invariante y el parámetro no-invariante, en el caso de las condiciones con 4 grupos) es inferior o igual al tamaño del efecto criterio, se considera que hay invarianza; mientras que si dicha diferencia es superior al tamaño del efecto criterio, se considera que no hay invarianza para el parámetro. Por tanto, en este caso la potencia es la proporción de réplicas en las que la diferencia entre los conjuntos es superior al tamaño del efecto criterio.

Se empleó un tamaño del efecto criterio distinto en función del tipo de parámetro sobre el cual se realizaba el contraste, aunque en ambos casos se utilizó un tamaño del efecto pequeño, basado en las diferencias estandarizadas entre los parámetros (Cohen, 1992; Pornprasertmanit, 2018). Así pues, en el contraste de invarianza en los pesos, se empleó un tamaño del efecto de .1, mientras que para el contraste en las intersecciones se utilizó un tamaño del efecto de .2.

Resultados

A continuación, se explican los resultados obtenidos respecto a las variables dependientes de interés: en el primer apartado, se comentan los resultados referentes al sesgo de las estimaciones y, en el segundo apartado, se habla sobre la potencia de los dos métodos analizados.

Sesgo de las estimaciones

Por motivos de espacio y simplicidad, en el cuerpo del trabajo se incluyen únicamente tablas (Tabla 1 a Tabla 4) con los promedios globales de RMSE para los parámetros más relevantes en cada caso. Los valores promedio globales de RMSE para el resto de los parámetros se incluyen en el Apéndice B (Tabla A2 a Tabla A5).

Antes de comentar los resultados de sesgo con más detalle, cabe destacar una serie de tendencias generales que se obtuvieron a partir de los resultados. En primer lugar, cada tipo de parámetro se estimó con diferente precisión. Así pues, tanto en el AFC-MG como en el método de alineamiento, las estimaciones de los pesos han sido las menos sesgadas, con los valores de RMSE más bajos, seguidas de las estimaciones de las intersecciones. En el AFC-MG, las varianzas son los parámetros estimados con mayor sesgo, con valores de RMSE más altos que las medias. Por su parte, en el método de alineamiento ocurre el patrón inverso, puesto que las medias se han estimado más sesgadas que las de las varianzas, salvo en un número limitado de condiciones extremas de no-invarianza en los pesos en las que los valores de RMSE para las varianzas han sido extremadamente altos.

En segundo lugar, es importante destacar también que, en todas las condiciones, independientemente del tipo de parámetro, el tipo de no-invarianza y el método de evaluación, el sesgo de las estimaciones disminuye a mayor tamaño muestral, de manera que los valores de RMSE han sido sistemáticamente menores cuando el tamaño muestral es igual a 1000 que cuando es de 500 y, sobre todo, 250. En tercer lugar, se ha observado que el tipo de no-invarianza tiene un efecto en los resultados de sesgo para ambos métodos, de manera que se observan tendencias distintas para cada tipo de parámetro según la localización de la no-invarianza.

A continuación, se explican más detalladamente los resultados para el AFC-MG y, más adelante, los del método de alineamiento.

Para el AFC-MG, se ha comprobado que en las condiciones de simulación en las que los pesos no son invariantes, se produce un aumento del sesgo en los pesos y las varianzas factoriales cuanto mayor es la magnitud de no-invarianza (Tablas 1 y 2), mientras que el sesgo de las intersecciones y de las medias se ha mantenido constante en valores de RMSE similares a los de invarianza completa, independientemente de los valores de los pesos no-invariantes. Sin embargo, en las condiciones en las que las intersecciones no son invariantes, se produce un aumento de los valores de RMSE a

mayor magnitud de no-invarianza en las intersecciones y las medias (Tablas 3 y 4), manteniéndose en valores constantes y similares a los de invarianza completa para los pesos y las varianzas factoriales. Por su parte, el número de grupos no parece tener un efecto sobre la precisión de las estimaciones de ninguno de los parámetros, ya que los valores de RMSE ni aumentan ni disminuyen de manera notable y consistente al incrementarse el número de grupos. Todas estas tendencias se pueden apreciar de manera gráfica en las Figuras 1 y 2 que aparecen a continuación.

Tabla 1
RMSE promedio de pesos para no-invarianza en pesos

| Método | Número de grupos | Tamaño muestral | $\Delta\lambda = 0$ | $\Delta\lambda = .1$ | $\Delta\lambda = .2$ | $\Delta\lambda = .3$ | $\Delta\lambda = .4$ |
|--------------|------------------|-----------------|---------------------|----------------------|----------------------|----------------------|----------------------|
| AFC-MG | 4 | 250 | .043 | .045 | .050 | .054 | .059 |
| | | 500 | .031 | .034 | .038 | .043 | .048 |
| | | 1000 | .022 | .026 | .030 | .035 | .040 |
| | 8 | 250 | .041 | .044 | .048 | .053 | .057 |
| | | 500 | .029 | .032 | .037 | .041 | .046 |
| | | 1000 | .020 | .024 | .029 | .034 | .039 |
| | 16 | 250 | .041 | .044 | .048 | .053 | .057 |
| | | 500 | .029 | .032 | .037 | .042 | .047 |
| | | 1000 | .020 | .024 | .029 | .034 | .039 |
| | 32 | 250 | .040 | .042 | .046 | .051 | .056 |
| | | 500 | .028 | .031 | .036 | .041 | .046 |
| | | 1000 | .020 | .023 | .029 | .034 | .039 |
| Alineamiento | 4 | 250 | .054 | .054 | .055 | .055 | .055 |
| | | 500 | .039 | .039 | .039 | .039 | .039 |
| | | 1000 | .027 | .027 | .028 | .027 | .027 |
| | 8 | 250 | .056 | .056 | .057 | .057 | .058 |
| | | 500 | .038 | .039 | .039 | .040 | .040 |
| | | 1000 | .027 | .027 | .028 | .028 | .028 |
| | 16 | 250 | .060 | .061 | .064 | .067 | .069 |
| | | 500 | .040 | .040 | .043 | .045 | .047 |
| | | 1000 | .028 | .029 | .030 | .032 | .033 |
| | 32 | 250 | .075 | .081 | .095 | .120 | .257 |
| | | 500 | .046 | .050 | .058 | .066 | .073 |
| | | 1000 | .030 | .032 | .039 | .045 | .050 |

Figura 1. RMSE promedio global para el AFC-MG en condiciones de no-invarianza en pesos en función de la magnitud de no-invarianza y el número de grupos

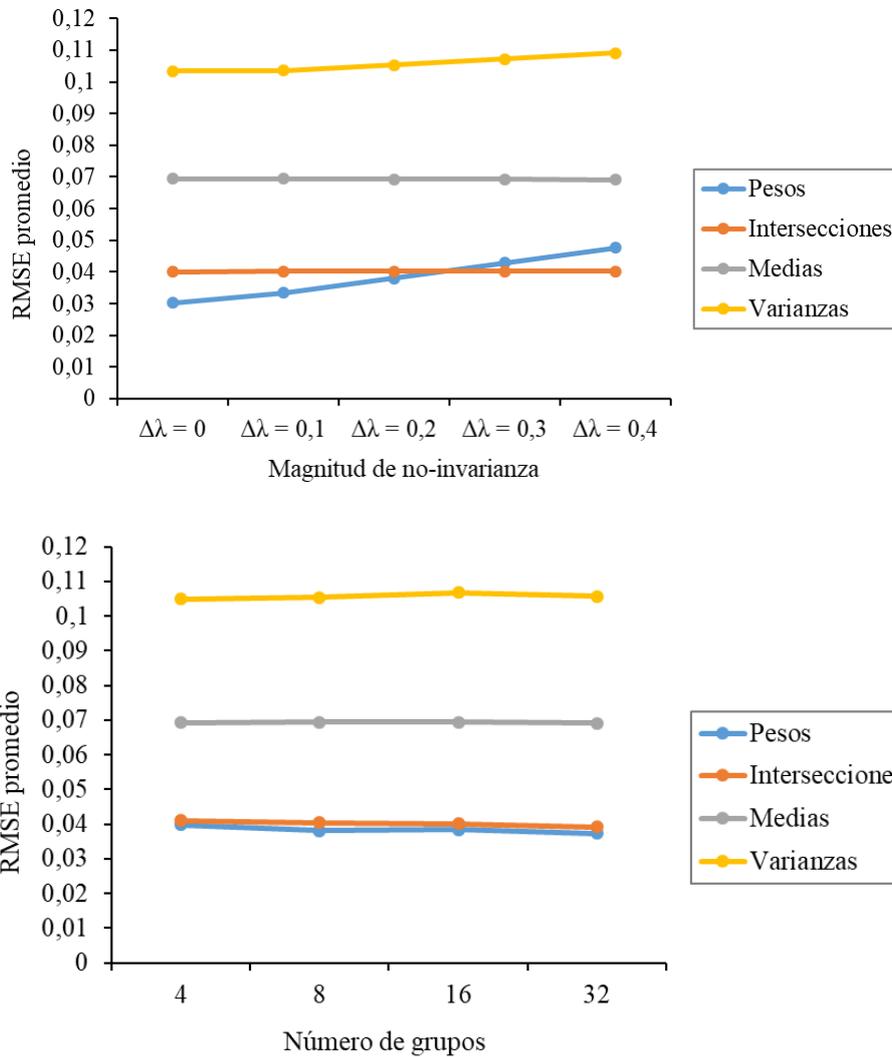
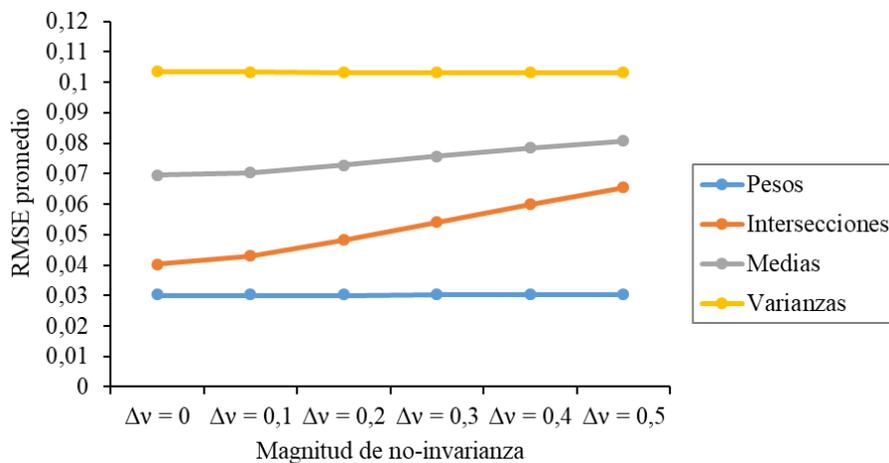


Figura 2. RMSE promedio global para el AFC-MG en condiciones de no-invarianza en intersecciones en función de la magnitud de no-invarianza y el número de grupos



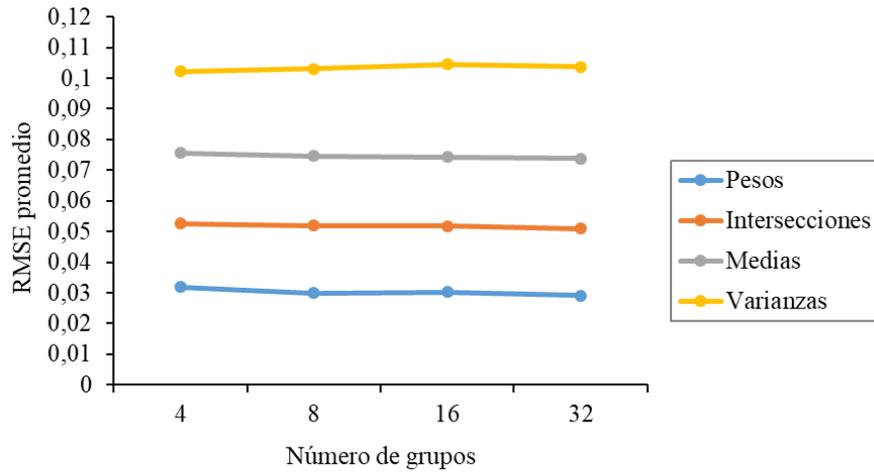


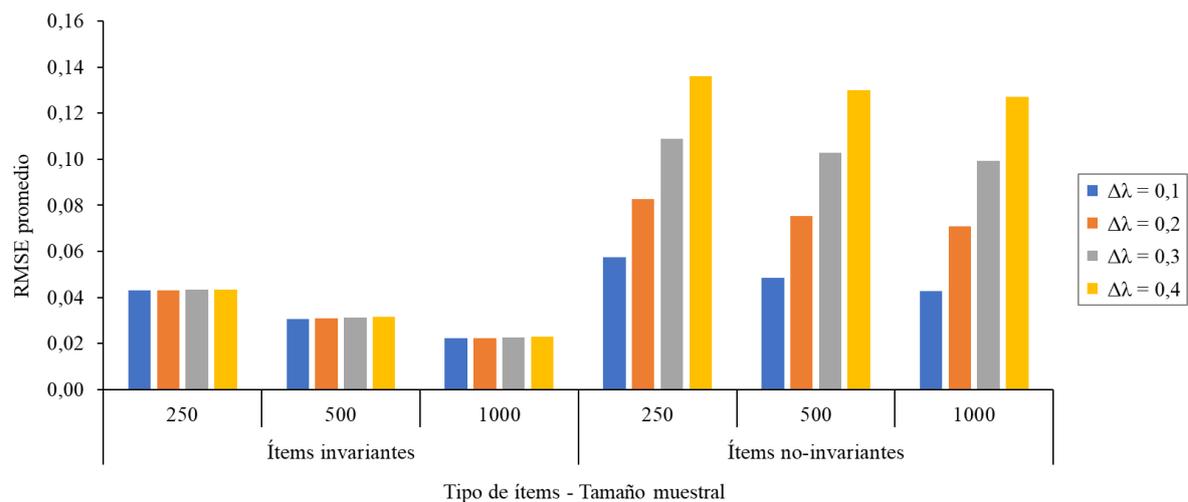
Tabla 2
RMSE promedio de varianzas para no-invarianza en pesos

| Método | Número de grupos | Tamaño muestral | $\Delta\lambda = 0$ | $\Delta\lambda = .1$ | $\Delta\lambda = .2$ | $\Delta\lambda = .3$ | $\Delta\lambda = .4$ |
|--------------|------------------|-----------------|---------------------|----------------------|----------------------|----------------------|----------------------|
| AFC-MG | 4 | 250 | .139 | .138 | .139 | .140 | .142 |
| | | 500 | .099 | .099 | .101 | .103 | .106 |
| | | 1000 | .069 | .070 | .073 | .077 | .079 |
| | 8 | 250 | .141 | .140 | .141 | .142 | .143 |
| | | 500 | .099 | .100 | .101 | .103 | .105 |
| | | 1000 | .069 | .070 | .073 | .076 | .078 |
| | 16 | 250 | .144 | .143 | .144 | .145 | .146 |
| | | 500 | .100 | .101 | .102 | .104 | .106 |
| | | 1000 | .069 | .071 | .073 | .076 | .078 |
| | 32 | 250 | .142 | .141 | .142 | .143 | .144 |
| | | 500 | .099 | .099 | .101 | .103 | .105 |
| | | 1000 | .070 | .071 | .073 | .076 | .078 |
| Alineamiento | 4 | 250 | .072 | .072 | .072 | .073 | .073 |
| | | 500 | .050 | .050 | .051 | .051 | .051 |
| | | 1000 | .034 | .035 | .035 | .035 | .035 |
| | 8 | 250 | .074 | .074 | .075 | .077 | .078 |
| | | 500 | .050 | .050 | .051 | .052 | .053 |
| | | 1000 | .035 | .036 | .036 | .037 | .037 |
| | 16 | 250 | .082 | .084 | .089 | .094 | .098 |
| | | 500 | .052 | .054 | .057 | .060 | .063 |
| | | 1000 | .036 | .037 | .040 | .042 | .045 |
| | 32 | 250 | .106 | .115 | .140 | .269 | 1.647 |
| | | 500 | .062 | .068 | .081 | .094 | .106 |
| | | 1000 | .039 | .043 | .053 | .061 | .070 |

Al examinar en detalle el sesgo según el tipo de ítem y grupo, se observa que los parámetros que se ven más afectados son los de los ítems y los grupos no-invariantes. Así pues, en las condiciones de no-invarianza en los pesos, las estimaciones de los pesos de

los ítems invariantes y de las varianzas de los grupos invariantes muestran valores de RMSE similares a los de invarianza completa; mientras que los pesos de los ítems no-invariantes y las varianzas de los grupos no-invariantes presentan mayores niveles de sesgo, con valores de RMSE más altos a menor tamaño muestral y, sobre todo, a mayor magnitud de no-invarianza. Por su parte, en las condiciones de no-invarianza en las intersecciones se observa un patrón similar, pero para las intersecciones y las medias. En la Figura 3 se pueden apreciar claramente estas tendencias (aunque se incluyen solamente los resultados de los pesos para las condiciones de no-invarianza en pesos con 4 grupos, el resto de condiciones y parámetros sigue un patrón similar). En ambos casos, no se observan diferencias en los valores de RMSE en función del tipo de ítem o grupo para los parámetros que no se ven afectados por el tipo de no-invarianza.

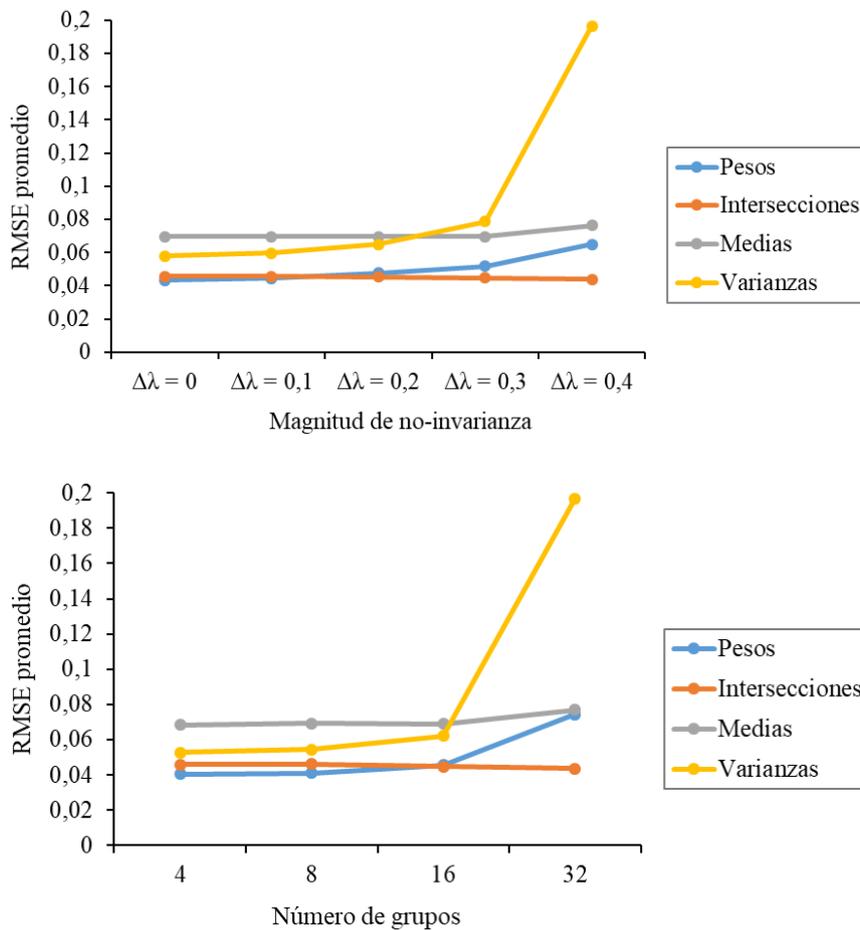
Figura 3. RMSE promedio para los pesos en el AFC-MG en condiciones de no-invarianza en pesos y 4 grupos en función del tipo de ítem



Para el método de alineamiento, se ha comprobado que la falta de invarianza en los pesos, al igual que ocurre en el AFC-MG, produce un aumento del sesgo en los pesos y las varianzas factoriales a mayor magnitud de no-invarianza (Tablas 1 y 2), mientras que el sesgo de las intersecciones y de las medias se mantiene constante en valores de RMSE similares a los de invarianza completa, tal y como se puede observar en la Figura 4. Sin embargo, el efecto de la magnitud de no-invarianza es mayor a menor muestra y mayor número de grupos, de manera que en las condiciones con menos grupos (esto es, 4 y 8 grupos) y tamaño muestral medio o alto apenas se produce un aumento de los valores de RMSE. Por tanto, el efecto de la magnitud de no-invarianza es, en general, menor que en el AFC-MG.

Por otro lado, para el alineamiento sí hay una influencia del número de grupos, de manera que se obtienen valores mayores de RMSE para los pesos y las varianzas a mayor número de grupos (Tablas 1 y 2), pero no para las intersecciones y las medias. Esta tendencia se puede apreciar en la Figura 4. Este efecto del número de grupos es más pronunciado cuanto menor es el tamaño muestral. De hecho, en determinadas condiciones extremas, con 32 grupos, tamaño muestral pequeño y magnitud de no-invarianza grande (esto es, $\Delta\lambda = .3$ y, sobre todo, $\Delta\lambda = .4$), RMSE toma valores muy altos para los pesos y, especialmente, para las varianzas (incluso las medias se ven afectadas).

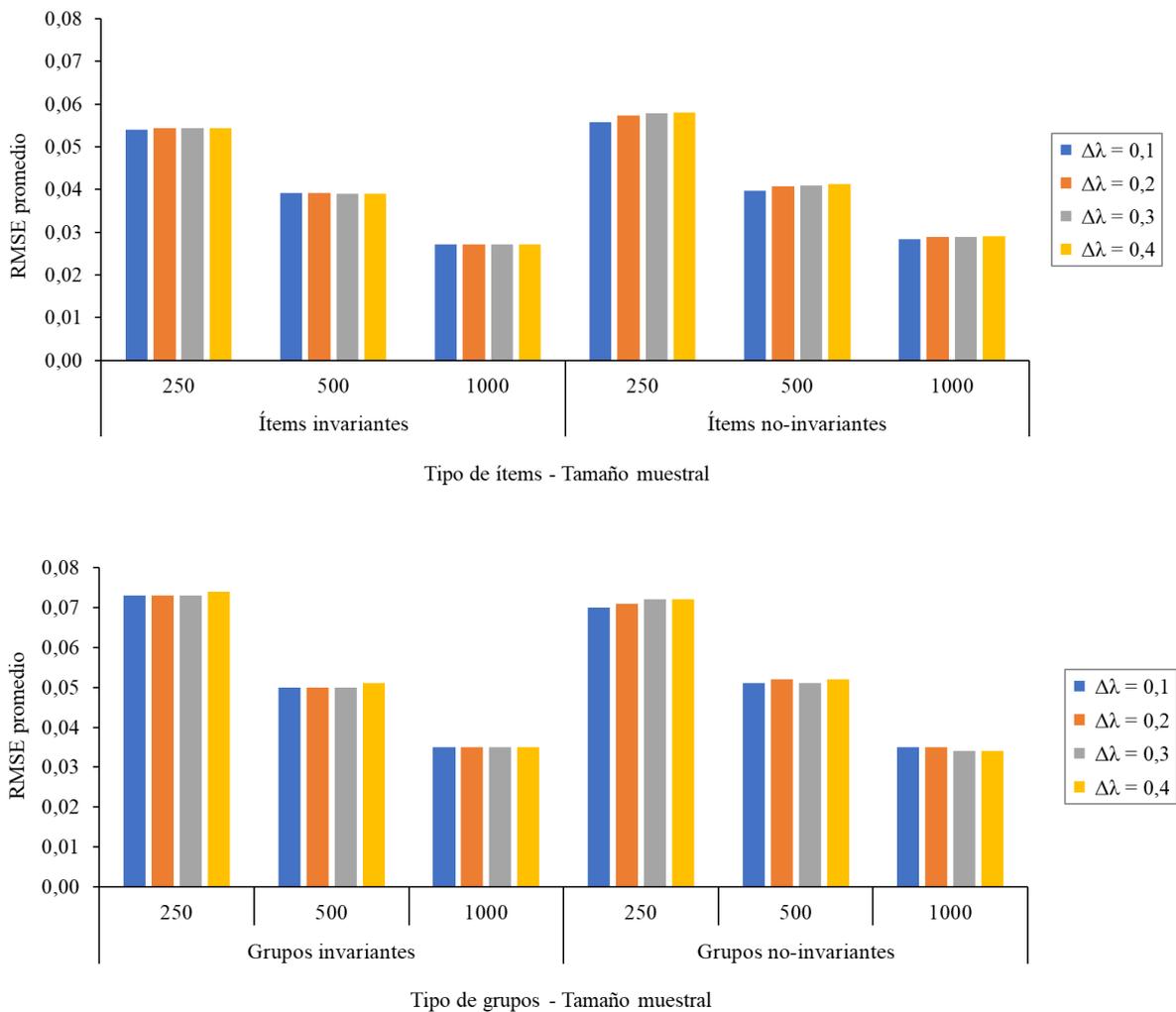
Figura 4. RMSE promedio global para el alineamiento en condiciones de no-invarianza en pesos en función de la magnitud de no-invarianza y el número de grupos



Al examinar los valores de RMSE según el tipo de ítem y grupo, se observa que la diferencia entre el sesgo de los parámetros (pesos y varianzas) invariantes y los no-invariantes es menor que en el AFC-MG, pareciéndose ambos al valor promedio global de RMSE para la condición concreta. No obstante, la diferencia entre ambos es menor cuando la muestra es grande que cuando es pequeña. Para los pesos, se observa que los

parámetros de los ítems invariantes se estiman algo mejor que los de los ítems no-invariantes, salvo con 32 grupos, donde parece que los pesos no-invariantes son estimados de forma más precisa. En el caso de las varianzas, ocurre el patrón contrario a los pesos, ya que los valores de RMSE son ligeramente menores para las varianzas de los grupos no-invariantes que para los grupos invariantes. En la Figura 5 se muestran gráficamente ambas tendencias (aunque se incluyen los resultados para las condiciones con 4 grupos, el resto de condiciones grupales siguen un patrón similar, salvo para la condición de 32 grupos en los pesos). Por su parte, no se observan diferencias en los valores de RMSE en función del tipo de ítem o grupo para las intersecciones y las medias, al igual que ocurre en el AFC-MG.

Figura 5. RMSE promedio para los pesos y las varianzas en el alineamiento en condiciones de no-invarianza en pesos y 4 grupos en función del tipo de ítem y grupo



Cuando la falta de invarianza se da en las intersecciones, se observa que el aumento de la magnitud de no-invarianza no conlleva un aumento de los valores de RMSE para ninguno de los parámetros, ni siquiera para las intersecciones o las medias (Tablas 3 y 4), al contrario de lo que ocurre en el AFC-MG. Por su parte, el número de grupos también hace que los valores de RMSE sean mayores, pero para los pesos y las varianzas, no para las intersecciones y las medias, donde los valores de RMSE se mantienen bastante constantes independientemente del número de grupos. Ambas tendencias pueden apreciarse en la Figura 6. Del mismo modo que ocurre en la no-invarianza en pesos, el efecto del número de grupos es mayor cuanto menor es el tamaño muestral.

Figura 6. RMSE promedio global para el alineamiento en condiciones de no-invarianza en intersecciones en función de la magnitud de no-invarianza y el número de grupos

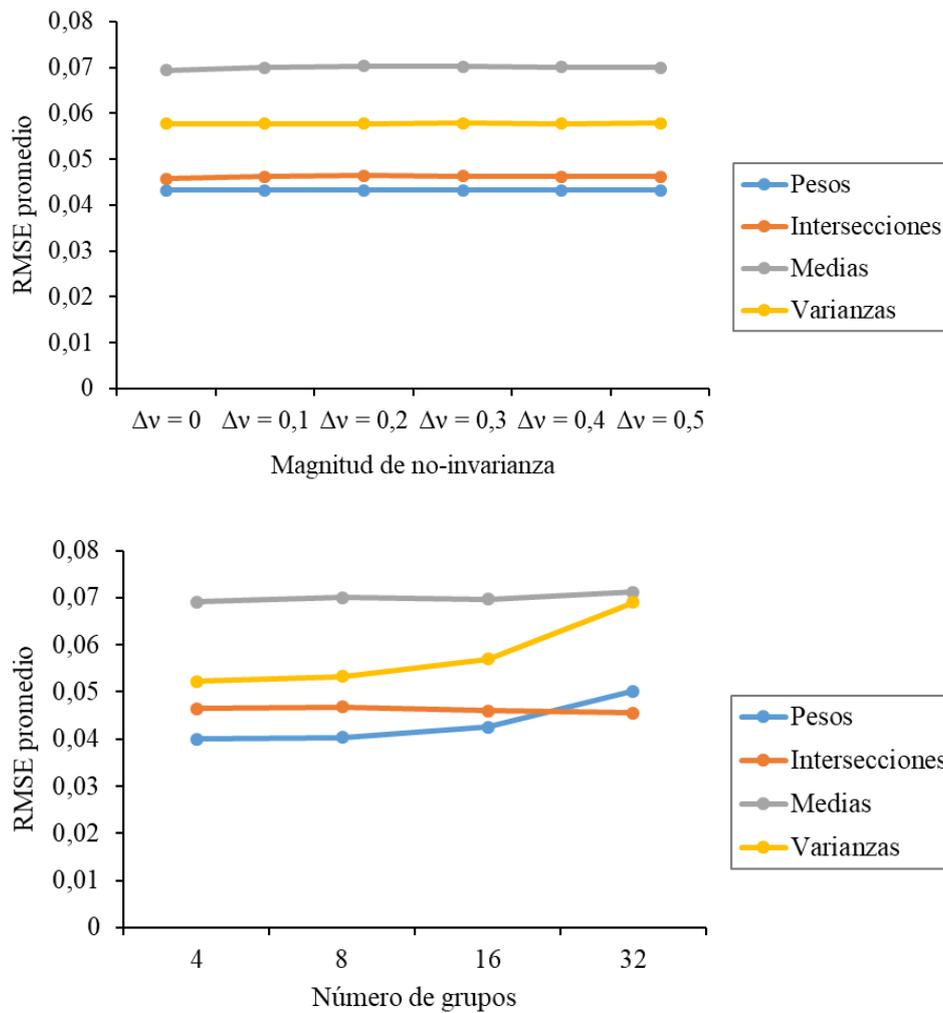


Tabla 3

RMSE promedio de intersecciones para no-invarianza en intersecciones

| Método | Número de grupos | Tamaño muestral | $\Delta v = 0$ | $\Delta v = .1$ | $\Delta v = .2$ | $\Delta v = .3$ | $\Delta v = .4$ | $\Delta v = .5$ |
|--------------|------------------|-----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| AFC-MG | 4 | 250 | .056 | .058 | .062 | .068 | .073 | .078 |
| | | 500 | .040 | .043 | .048 | .053 | .059 | .065 |
| | | 1000 | .028 | .031 | .037 | .043 | .050 | .055 |
| | 8 | 250 | .055 | .057 | .062 | .067 | .072 | .077 |
| | | 500 | .039 | .042 | .047 | .053 | .059 | .065 |
| | | 1000 | .027 | .031 | .037 | .043 | .049 | .055 |
| | 16 | 250 | .055 | .057 | .061 | .066 | .072 | .077 |
| | | 500 | .039 | .042 | .047 | .053 | .059 | .064 |
| | | 1000 | .027 | .031 | .036 | .043 | .049 | .055 |
| | 32 | 250 | .053 | .055 | .060 | .066 | .071 | .076 |
| | | 500 | .039 | .041 | .047 | .052 | .058 | .064 |
| | | 1000 | .026 | .030 | .036 | .042 | .048 | .054 |
| Alineamiento | 4 | 250 | .063 | .064 | .064 | .064 | .064 | .064 |
| | | 500 | .044 | .045 | .045 | .045 | .045 | .045 |
| | | 1000 | .031 | .031 | .031 | .031 | .031 | .031 |
| | 8 | 250 | .063 | .064 | .064 | .064 | .064 | .064 |
| | | 500 | .045 | .045 | .045 | .045 | .045 | .045 |
| | | 1000 | .031 | .032 | .032 | .032 | .032 | .032 |
| | 16 | 250 | .062 | .062 | .063 | .063 | .062 | .062 |
| | | 500 | .043 | .043 | .044 | .044 | .043 | .043 |
| | | 1000 | .032 | .032 | .032 | .032 | .032 | .032 |
| | 32 | 250 | .061 | .061 | .062 | .062 | .061 | .061 |
| | | 500 | .044 | .044 | .044 | .044 | .044 | .044 |
| | | 1000 | .031 | .032 | .032 | .032 | .031 | .031 |

Al examinar el sesgo según el tipo de ítem y grupo, se observa que, de nuevo, los valores de ambos tipos para las intersecciones y las medias son parecidos al RMSE promedio global, siendo la diferencia entre ellos mucho menor que en el AFC-MG y en la condición de no-invarianza en los pesos. A pesar de que parece que se estiman ligeramente mejor los parámetros de los ítems y los grupos invariantes, la diferencia entre ambos es muy pequeña. Por su parte, no se observan diferencias en los valores de sesgo en función del tipo de ítem o grupo para los pesos y las varianzas, al igual que en el AFC-MG. Estos resultados se pueden apreciar gráficamente en la Figura 7 (se incluyen únicamente los resultados para las intersecciones en las condiciones con 4 grupos, pero para el resto de parámetros y condiciones grupales la tendencia es muy similar).

Figura 7. RMSE promedio para las intersecciones en el alineamiento en condiciones de no-invarianza en intersecciones y 4 grupos en función del tipo de ítem y grupo

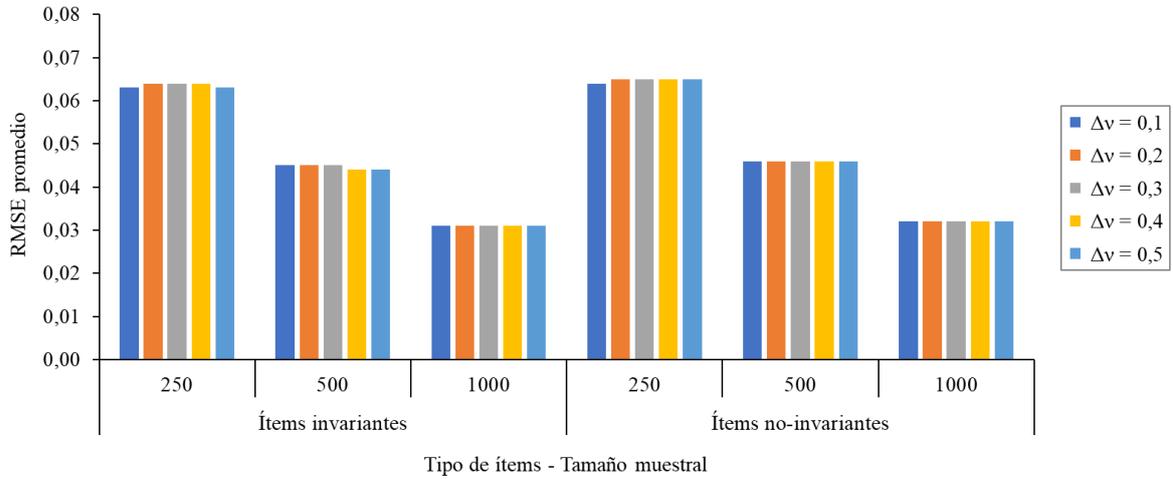


Tabla 4
RMSE promedio de medias para no-invarianza en intersecciones

| Método | Número de grupos | Tamaño muestral | $\Delta v = 0$ | $\Delta v = .1$ | $\Delta v = .2$ | $\Delta v = .3$ | $\Delta v = .4$ | $\Delta v = .5$ |
|--------------|------------------|-----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| AFC-MG | 4 | 250 | .094 | .095 | .097 | .100 | .103 | .106 |
| | | 500 | .067 | .068 | .071 | .074 | .077 | .080 |
| | | 1000 | .047 | .049 | .052 | .057 | .060 | .063 |
| | 8 | 250 | .095 | .096 | .097 | .100 | .102 | .104 |
| | | 500 | .067 | .068 | .070 | .073 | .076 | .078 |
| | | 1000 | .046 | .048 | .051 | .054 | .058 | .060 |
| | 16 | 250 | .095 | .096 | .097 | .099 | .101 | .103 |
| | | 500 | .067 | .068 | .070 | .073 | .076 | .078 |
| | | 1000 | .047 | .048 | .051 | .054 | .057 | .059 |
| | 32 | 250 | .094 | .095 | .096 | .099 | .101 | .103 |
| | | 500 | .067 | .068 | .070 | .072 | .075 | .075 |
| | | 1000 | .047 | .048 | .051 | .054 | .057 | .059 |
| Alineamiento | 4 | 250 | .095 | .096 | .096 | .096 | .096 | .096 |
| | | 500 | .065 | .065 | .066 | .066 | .065 | .065 |
| | | 1000 | .046 | .046 | .047 | .046 | .046 | .046 |
| | 8 | 250 | .095 | .096 | .096 | .096 | .096 | .096 |
| | | 500 | .066 | .067 | .067 | .067 | .067 | .067 |
| | | 1000 | .047 | .047 | .047 | .047 | .047 | .047 |
| | 16 | 250 | .095 | .096 | .096 | .096 | .096 | .096 |
| | | 500 | .065 | .066 | .066 | .066 | .066 | .066 |
| | | 1000 | .047 | .048 | .048 | .048 | .048 | .048 |
| | 32 | 250 | .098 | .099 | .099 | .099 | .099 | .099 |
| | | 500 | .067 | .068 | .068 | .068 | .067 | .067 |
| | | 1000 | .047 | .048 | .048 | .048 | .048 | .047 |

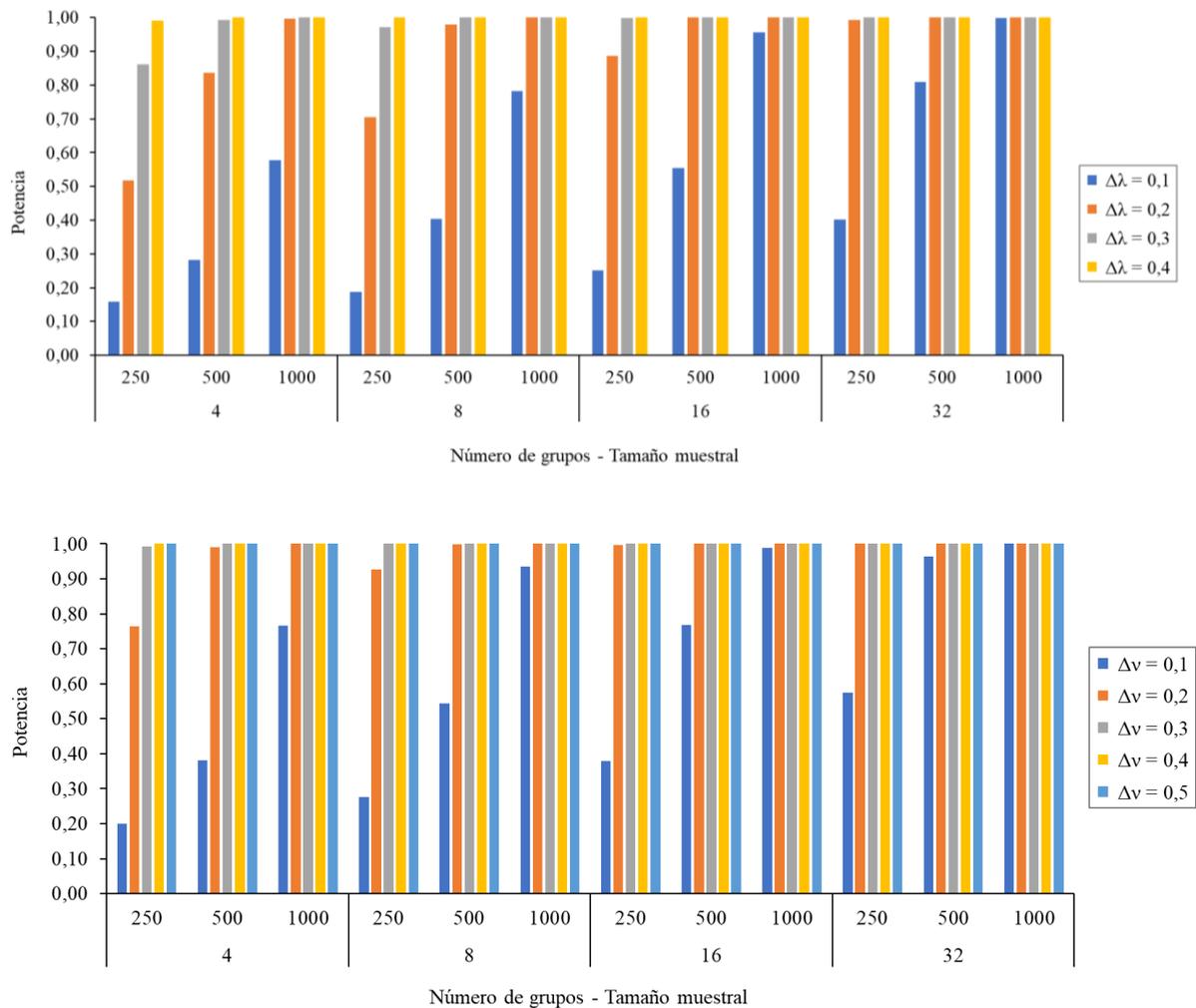
Por último, cabe destacar que la precisión de las estimaciones ha sido mejor en un método que en el otro, en función del tipo de parámetro o de no-invarianza. Así pues, los pesos han sido estimados de manera más precisa con el AFC-MG que con el método de alineamiento en todas las condiciones. En cambio, las intersecciones son mejor estimadas con el AFC-MG que con el alineamiento con falta de invarianza en los pesos, pero con no-invarianza en las intersecciones se han estimado de forma más precisa con el alineamiento (salvo para invarianza completa y $\Delta v = .1$). Las medias se han estimado de manera similar con ambos métodos con no-invarianza en los pesos, pero el alineamiento ha realizado mejores estimaciones que el AFC-MG cuando la falta de invarianza está en las intersecciones, especialmente a mayor magnitud de no-invarianza. Por su parte, las estimaciones de las varianzas han sido más precisas con el método de alineamiento que con el AFC-MG, salvo en algunas condiciones extremas (es decir, con 32 grupos, tamaño muestral pequeño y magnitud de no-invarianza grande).

Potencia del método

Antes de explicar detalladamente los resultados de potencia de ambos métodos (las tablas con los resultados se adjuntan en el Apéndice C), es necesario comentar que no se pretende comparar los resultados de ambas técnicas directamente, ya que el procedimiento de detección de no-invarianza es distinto en cada método: en el AFC-MG es un procedimiento de contraste de hipótesis basado en la significación estadística; mientras que en el alineamiento el procedimiento se basa en tamaños del efecto.

Respecto al AFC-MG, se puede observar que, tanto en el contraste de la invarianza métrica como en el de la invarianza escalar, la potencia aumenta a mayor tamaño muestral, mayor número de grupos y mayor magnitud de no-invarianza. Sin embargo, el efecto del número de grupos es más notable en las condiciones con muestra pequeña ($n = 250$) y magnitud de no-invarianza pequeña ($\Delta = .1$). En estas condiciones, la potencia es baja y, por tanto, se puede observar claramente un incremento de la potencia al aumentar los grupos. En condiciones de magnitud y muestra grande, la potencia es de por sí tan alta (en torno a uno) que el efecto de los grupos no se aprecia. Dichas tendencias se pueden observar en las gráficas de la Figura 8.

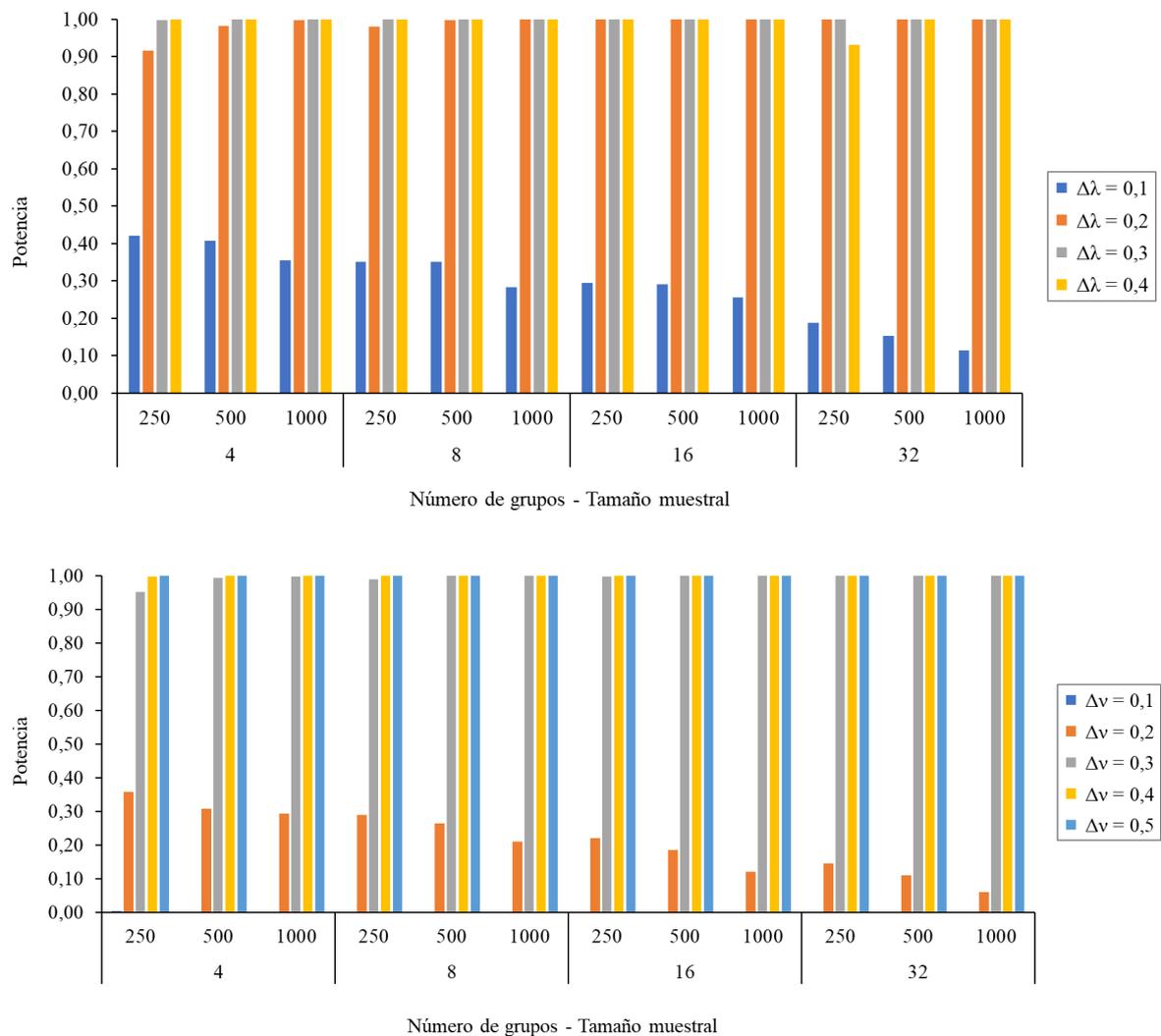
Figura 8. Potencia del AFC-MG para la invarianza métrica y escalar en función de la magnitud de no-invarianza, el número de grupos y el tamaño muestral



Por su parte, en el método de alineamiento se observa que para la detección de no-invarianza tanto en los pesos como en las intersecciones la potencia aumenta al incrementarse la magnitud de no-invarianza. El tamaño muestral parece influir de manera distinta según la condición: cuando la magnitud de no-invarianza coincide con el tamaño del efecto límite para el contraste ($\Delta = .1$ para pesos, $\Delta = .2$ para intersecciones), la potencia es ligeramente menor a mayor tamaño muestral; mientras que en las condiciones en las que la magnitud de no-invarianza es superior a dicho tamaño del efecto, la potencia aumenta. Respecto al número de grupos, parece ocurrir algo similar: si la magnitud de no-invarianza es igual al tamaño del efecto límite, la potencia del procedimiento disminuye a mayor número de grupos, pero no parece ser el caso cuando la magnitud de no-invarianza es superior a dicho tamaño del efecto.

Al igual que ocurre en el AFC-MG, estos efectos son más notables en las condiciones en las que la potencia es baja (esto es, con 4 grupos, muestra pequeña y, sobre todo, magnitud de no-invarianza pequeña), ya que en la mayoría de los casos en los que la magnitud de no-invarianza es superior al tamaño del efecto límite, la potencia es de uno. Estas tendencias se pueden apreciar de manera gráfica en la Figura 9. Por último, cabe destacar que para el contraste de invarianza métrica en el alineamiento en la condición más extrema (es decir, 32 grupos, tamaño muestral pequeño y magnitud de no-invarianza grande) se produce un ligero descenso de la potencia.

Figura 9. Potencia del alineamiento para la invarianza métrica y escalar en función de la magnitud de no-invarianza, el número de grupos y el tamaño muestral



Discusión y conclusiones

El propósito de este trabajo ha consistido en investigar el funcionamiento y la calidad de las estimaciones de dos métodos de prueba de invarianza distintos: el AFC-MG tradicional y el recientemente propuesto método de alineamiento. Concretamente, el objetivo era estudiar la precisión de sus estimaciones de los parámetros, así como su potencia a la hora de detectar la falta de invarianza en función de cuatro factores: el tipo de no-invarianza, la magnitud de dicha no-invarianza, el número de grupos y el tamaño muestral grupal. Además, un objetivo adicional fue proponer un procedimiento de contraste de invarianza para el alineamiento basado en tamaños del efecto en lugar de en significación estadística. Las hipótesis de partida fueron que (1) el sesgo aumentaría a mayor magnitud de no-invarianza, mayor número de grupos y menor tamaño muestral en ambos métodos; (2) el sesgo se vería afectado por la localización de la no-invarianza en ambos métodos; (3) el método de alineamiento estimaría de forma más precisa los parámetros que el AFC-MG; (4) la potencia del AFC-MG aumentaría a mayor magnitud de no-invarianza, mayor número de grupos y mayor tamaño muestral; y (5) la potencia del alineamiento aumentaría a mayor magnitud de no-invarianza, pero sería independiente del tamaño muestral.

En primer lugar, se ha observado que, tal y como se había hipotetizado en un principio, las estimaciones de todos los parámetros estuvieron menos sesgadas cuanto mayor fue el tamaño muestral de los grupos para ambos métodos. Esto va en la línea de los estudios previos, que también han comprobado que las estimaciones resultan más precisas a mayor muestra para el alineamiento y el AFC-MG (Asparouhov y Muthén, 2014; Marsh *et al.*, 2018; Meade y Bauer, 2007). Este resultado era esperable, ya que las estimaciones con ML tienen propiedades asintóticas cuando el tamaño muestral es alto (Reinartz, Haenlein y Henseler, 2009).

En segundo lugar, también en la línea de lo esperado, se ha observado que el tipo de no-invarianza influye de forma diferente en el sesgo de los distintos parámetros. Para el AFC-MG, se ha podido comprobar que, cuando los pesos no son invariantes, se ve afectada la precisión de los pesos y las varianzas factoriales; mientras que en las condiciones en las que la no-invarianza se localiza en las intersecciones, se producen variaciones en la precisión de las estimaciones de las intersecciones y las medias. En general, estos resultados van en la línea de otros estudios (Chen, 2008) y de las bases teóricas de la invarianza: para que las comparaciones en las varianzas sean válidas, es

necesario que los pesos sean iguales; mientras que, para poder comparar las medias, las intersecciones tienen que ser similares (Meredith y Teresi, 2006; Steenkamp y Baumgartner, 1998).

No obstante, la teoría plantea que para la comparación válida de medias grupales también es necesaria la igualdad en los pesos factoriales. Sin embargo, los resultados obtenidos no muestran que el aumento de la magnitud de las diferencias en los pesos tenga un impacto sobre el sesgo de las medias. Este resultado es distinto al que obtuvo Chen (2008), que mostró que la falta de invarianza en pesos producía un aumento del sesgo en las medias. Esta discordancia se puede deber a que la autora estudió el efecto de la cantidad de no-invarianza (es decir, el número de ítems no-invariantes), en lugar de su magnitud (es decir, el valor de las diferencias entre los parámetros). En este trabajo, la cantidad de no-invarianza se ha mantenido constante (16,67% de ítems, 25% de grupos) en valores inferiores a los estudiados por Chen (2008). Así pues, podría ser que el sesgo de las medias se viese más afectado por la cantidad de no-invarianza que por su magnitud, por lo que sería interesante investigar ambos efectos de manera conjunta en el futuro para comprobar si tienen un efecto distinto. Por otro lado, puede deberse al impacto diferente que tiene cada parámetro en la estimación de las medias. Steinmetz (2013) también observó que la falta de invarianza en los pesos no tenía un gran impacto en la precisión de estimación de las diferencias de medias observadas, y argumentó que podría deberse a que las intersecciones tienen un efecto mayor que los pesos sobre las medias: tal y como se planteó en la Ecuación 2, las intersecciones son constantes aditivas a las medias observadas, mientras que la influencia de los pesos está mediada por el factor latente, al cual multiplica (Steinmetz, 2013).

Por su parte, los resultados de sesgo para el alineamiento solo se han visto afectados por el tipo de no-invarianza en el caso de la falta de invarianza en pesos, produciendo un impacto en la precisión de los pesos y las varianzas. En cambio, la no-invarianza en intersecciones no ha influido en la precisión de ninguno de los parámetros, ni siquiera en la de las intersecciones o las medias factoriales. Estos resultados no se corresponden completamente con los del único estudio que ha investigado el efecto del tipo de no-invarianza. Flake y McCoach (2018) también observaron que la falta de invarianza en los pesos afectaba a la precisión en los pesos y las varianzas, pero la no-invarianza en los umbrales tuvo un impacto en el sesgo de los umbrales y las medias. Esta diferencia en los resultados puede deberse al tipo de ítems estudiados. En este trabajo, se

han empleado ítems continuos, al igual que en la mayoría de estudios de simulación del alineamiento, mientras que las autoras emplearon ítems politómicos con cinco categorías. La estructura de medias del modelo factorial es más compleja con ítems politómicos que con ítems continuos, tanto más cuantas más categorías de respuesta tengan los ítems (Flake y McCoach, 2018), por lo que es posible que el alineamiento funcione peor con ítems politómicos, que implican muchos más parámetros de medias, que con ítems continuos, donde parece que el alineamiento es capaz de realizar estimaciones bastante precisas de las intersecciones y las medias factoriales, al menos en las condiciones estudiadas. Así pues, sería interesante investigar en el futuro si la complejidad introducida por emplear ítems politómicos empeora los resultados de las medias frente a lo obtenido con ítems continuos.

En tercer lugar, se ha observado que la magnitud de no-invarianza ha producido un aumento del sesgo de las estimaciones de los parámetros afectados por el tipo de no-invarianza para ambas técnicas, en línea con las hipótesis de partida y con los resultados de estudios previos (Finch, 2016; Flake y McCoach, 2018; Marsh *et al.*, 2018). Así pues, parece que, en general, ambas técnicas (especialmente, el AFC-MG) realizan mejores estimaciones con magnitudes pequeñas de diferencias en los parámetros que con magnitudes grandes.

No obstante, cabe destacar que el efecto de esta variable no es el mismo para todos los parámetros en ambas técnicas. En el AFC-MG, este aumento de la magnitud de no-invarianza afecta principalmente a los parámetros de los ítems y los grupos no-invariantes, mientras que los de los elementos invariantes se estiman con la misma precisión que en las condiciones de invarianza completa. Este resultado era esperable: el AFC-MG impone la restricción de igualdad a todos los parámetros (pesos e intersecciones) a la hora de estimar el modelo de invarianza escalar. En este trabajo, todos los pesos eran iguales salvo los del ítem 2, así como todas las intersecciones, salvo las del ítem 4. Así pues, cabía esperar que todos los pesos e intersecciones invariantes se estimaran de forma más precisa porque son, de hecho, iguales entre sí. Sin embargo, al imponer la restricción de igualdad en los parámetros no-invariantes, sus estimaciones se sesgan en la dirección de los parámetros invariantes, alejándose de su valor real. En el alineamiento, en cambio, se observó que tanto los parámetros de los elementos invariantes y los no-invariantes se estimaron de manera bastante similar entre sí, lo cual era esperable

porque, al contrario que en el AFC-MG, no se imponen restricciones de igualdad en la estimación de los mismos.

Por otro lado, el número de grupos tiene un impacto claro en el sesgo de las estimaciones únicamente en el caso del alineamiento, ya que se produce un incremento de la magnitud del sesgo a medida que aumenta el número de grupos, aunque solo para los pesos y las varianzas. En cambio, en el AFC-MG no parece que el número de grupos influya en la precisión de estimación de los parámetros. Por tanto, la hipótesis de partida solo se ve apoyada para el alineamiento, donde los resultados van en la línea de estudios previos que han señalado que un aumento del número de grupos puede perjudicar la precisión de las estimaciones del alineamiento, especialmente en condiciones con muestras pequeñas y una gran cantidad de no-invarianza (Asparouhov y Muthén, 2014; Flake y McCoach, 2018), como ha ocurrido en este trabajo. El hecho de que no haya habido influencia del número de grupos en el AFC-MG puede deberse a que, al estimar el modelo de invarianza escalar, los parámetros (pesos e intersecciones) se estiman libremente solo para el primer grupo, que sería el grupo de referencia, tomando el resto de los parámetros los mismos valores que en este grupo debido a las restricciones de igualdad impuestas por el modelo (Byrne y van de Vijver, 2010). En cualquier caso, sería interesante comprobar si se replican estos resultados en otros estudios.

Respecto al método que da lugar a estimaciones más precisas, se ha comprobado que los resultados dependen del tipo de parámetro, por lo que la hipótesis de partida no se ha visto apoyada totalmente. Así pues, parece que los pesos están consistentemente menos sesgados en el AFC-MG, mientras que la precisión de las intersecciones depende del tipo de no-invarianza. Sin embargo, lo más relevante es que las estimaciones de las medias y las varianzas han sido igual o más precisas en el alineamiento que en el AFC-MG, salvo para ciertas condiciones extremas con muchos grupos, muestra pequeña y magnitud de no-invarianza grande. Estos resultados que apuntan a una estimación más precisa de las medias en el alineamiento coinciden con los de otros estudios que han investigado la calidad de las estimaciones de ambos métodos en condiciones similares a las de este trabajo (Marsh *et al.*, 2018; Pokropek *et al.*, 2019). Además, cabe destacar que estos resultados apoyan los principios del método de alineamiento: el procedimiento permite estimar las medias y varianzas factoriales de los grupos a la vez que minimiza la no-invarianza en los pesos y las intersecciones (Asparouhov y Muthén, 2014), bajo la asunción de invarianza aproximada.

Los resultados de potencia del AFC-MG han apoyado la hipótesis inicial, ya que se ha observado un incremento de la potencia para el contraste tanto de invarianza métrica como escalar a mayor tamaño muestral, magnitud de no-invarianza y número de grupos. No obstante, lo más relevante es el hecho de que, incluso cuando la magnitud de las diferencias es pequeña, parece que el aumento del número de grupos incrementa la potencia. Además, estos resultados van en la línea de estudios previos (Kim *et al.*, 2017; Meade y Bauer, 2007; Rutkowski y Svetina, 2014). Por su parte, la potencia del alineamiento utilizando tamaños del efecto ha incrementado a mayor magnitud de no invarianza, tal y como se esperaba. No obstante, en contra de la hipótesis inicial, no ha sido independiente del tamaño muestral, así como tampoco lo ha sido del número de grupos: se ha observado que en magnitudes de no-invarianza iguales a los tamaños del efecto límite utilizados para el contraste, la potencia disminuía a más grupos y tamaño muestral. Estos resultados podrían deberse, por un lado, a que hubiera una tendencia a la infraestimación de parámetros por parte del alineamiento y, por otro lado, a un incremento de la eficiencia de las estimaciones a mayor tamaño muestral (Reinartz *et al.*, 2009).

A partir de los resultados obtenidos, se pueden extraer una serie de conclusiones. Primero, teniendo en cuenta los resultados respecto al sesgo de las estimaciones con ambos métodos, parece que lo más recomendable es emplear muestras grandes si se desean estimaciones más precisas de los parámetros del modelo, sobre todo para las medias y las varianzas factoriales de los grupos y cuando se espera que las diferencias entre los parámetros de los grupos sean grandes.

Además, cuando se trabaja con el método de alineamiento con muchos grupos, y a riesgo de que las magnitudes de las diferencias en los parámetros puedan ser grandes, lo más recomendable es trabajar con tamaños muestrales grandes para que las estimaciones de los parámetros sean precisas. Como ya ha sido apuntado por otros autores, como Muthén y Asparouhov (2013), el alineamiento parece funcionar adecuadamente con un número de grupos moderado y con tamaños muestrales medios o grandes. Sin embargo, con un número muy elevado de grupos (por ejemplo, 100) el método de alineamiento puede no resultar adecuado, incluso con tamaños muestrales grandes, ya que puede dar problemas de convergencia (Muthén y Asparouhov, 2013).

Al tener en cuenta también los resultados de potencia, parece que cada una de las técnicas puede resultar adecuada en función del objetivo del investigador. Cuando el interés está en comprobar si un instrumento de medida concreto es completamente

invariante, el AFC-MG resulta una técnica adecuada (Kim *et al.*, 2017), y trabajar con muchos grupos y tamaños muestrales grandes puede resultar favorable, ya que incluso diferencias pequeñas serán detectadas con una probabilidad más alta. En cambio, si lo que se desea es lidiar con esas diferencias pequeñas para realizar comparaciones grupales, el AFC-MG puede ser poco adecuado, como ya han comentado otros autores (Byrne y van de Vijver, 2010).

El alineamiento puede resultar en estos casos una técnica más adecuada por varios motivos. En primer lugar, trabaja sobre la asunción más realista de que existen diferencias pequeñas pero irrelevantes entre los parámetros de los grupos. Así pues, si se espera que existan estas diferencias entre los grupos estudiados, el alineamiento puede facilitar el proceso de puesta a prueba de invarianza y realizar estimaciones más precisas. De hecho, en condiciones similares a las estudiadas en este trabajo (es decir, pocos ítems y grupos no-invariantes) el alineamiento puede ser una técnica más adecuada que el AFC-MG si lo que interesa es obtener una estimación precisa de las medias y las varianzas factoriales para realizar comparaciones grupales. Por otro lado, el alineamiento lleva a cabo un procedimiento *ad-hoc* de manera automática que da información sobre qué grupos son no-invariantes en cada uno de los parámetros (Asparouhov y Muthén, 2014; Kim *et al.*, 2017; Muthén y Asparouhov, 2013), por lo que este método puede ser útil para investigar las causas de la falta de invarianza de un determinado instrumento de medida.

Sin embargo, el alineamiento no carece de ciertas limitaciones. En primer lugar, se ha señalado de manera consistente que es necesario que la asunción de invarianza aproximada (es decir, que la mayoría de las diferencias entre los parámetros de los grupos son pequeñas e irrelevantes) se cumpla para que los resultados de la técnica sean adecuados (Asparouhov y Muthén, 2014; Kim *et al.*, 2017; Muthén y Asparouhov, 2013; Pokropek *et al.*, 2019). Aunque Muthén y Asparouhov (2014) propusieron que si la cantidad de no-invarianza es inferior al 25% de los parámetros los resultados pueden ser confiables, aún no parece estar claro cuánta falta de invarianza es realmente permisible para que los resultados del alineamiento sean válidos (Kim *et al.*, 2017). Además, el procedimiento *ad-hoc* ha sido criticado por ser demasiado conservador (Kim *et al.*, 2017; Flake y McCoach, 2018) y estar basado en la significación estadística (Marsh *et al.*, 2018).

Por otro lado, el alineamiento es menos flexible respecto a las especificaciones del modelo que el AFC-MG, ya que no permite incluir *cross-loadings* o covariables

(Asparouhov y Muthén, 2014; Kim *et al.*, 2017; Marsh *et al.*, 2018; Muthén y Asparouhov, 2013). Recientemente, Marsh *et al.* (2018) propusieron el *alignment-within-CFA* (AwC) para superar algunas de estas limitaciones y acercar el método de alineamiento a estrategias de carácter más confirmatorio. Sin embargo, el AwC es muy reciente, con muchos interrogantes acerca de su funcionamiento y calidad, y cuya base sigue siendo el alineamiento.

No obstante, el AFC-MG y el alineamiento no deben ser consideradas técnicas antitéticas, sino todo lo contrario. Dados los puntos fuertes y débiles de cada una de ellas, el alineamiento puede resultar adecuado en situaciones donde el AFC-MG resulta más limitado y viceversa. Además, el alineamiento puede entenderse como un método exploratorio (Asparouhov y Muthén, 2014) frente al carácter confirmatorio del AFC-MG, y pueden resultar útiles en distintas situaciones multigrupo en función de los objetivos y las expectativas de los investigadores, tal y como se ha comentado anteriormente.

Limitaciones y líneas futuras de investigación

Este trabajo presenta una serie de limitaciones. Primero, las condiciones de simulación han sido reducidas en comparación con la cantidad de variables que influyen en las aplicaciones reales. Por tanto, los resultados obtenidos y las conclusiones derivadas de estos deben generalizarse con precaución, limitándose únicamente a situaciones similares a las que se han estudiado aquí.

Además, para poder mantener el número de condiciones manejable, ha habido muchas variables de interés que no se han podido estudiar. Por ejemplo, se ha trabajado en condiciones bastante favorables para ambas técnicas, y hubiera resultado interesante estudiar el efecto de la cantidad de no-invarianza no solo para comprobar si su influencia en los resultados es consistente con la de otros estudios, sino para arrojar luz sobre el debate de cuánta no-invarianza es suficientemente pequeña para que el alineamiento proporcione resultados adecuados. Tampoco se ha estudiado la asimetría de los datos: a pesar de que se ha dicho que el alineamiento permite cualquier tipo de distribución (Muthén y Asparouhov, 2013), se ha señalado que la asimetría podría influir negativamente en la precisión de las estimaciones y la potencia de la técnica (Flake y McCoach, 2018), por lo que sería interesante investigar en el futuro el efecto de la asimetría sobre los resultados del alineamiento.

Del mismo modo, tampoco se han estudiado los métodos bajo condiciones más realistas, como la de ítems politómicos. Se emplearon ítems continuos en este trabajo para ser consistente con lo que han hecho otros autores y poder comparar con sus resultados. No obstante, los ítems que se emplean en las aplicaciones reales son politómicos, tal y como se comprobó en la revisión de la literatura, por lo que es posible que los resultados obtenidos no se correspondan totalmente con la práctica aplicada. Por tanto, sería interesante estudiar el funcionamiento del alineamiento con ítems politómicos. Asimismo, tampoco se han considerado situaciones con más de una variable latente. Aunque la mayoría de los estudios aplicados revisados emplearon medidas de un factor, no todos los constructos psicológicos son unidimensionales. De hecho, hubo bastantes trabajos con instrumentos que medían más de una variable latente, por lo que sería interesante comprobar el funcionamiento de la técnica para modelos multidimensionales.

Por otro lado, el procedimiento de potencia basado en tamaños del efecto se ha propuesto como una alternativa que puede superar algunas de las críticas que ha recibido el alineamiento. No obstante, hay que tener en cuenta varias cosas. En primer lugar, en este trabajo se ha llevado a cabo el procedimiento de manera simplificada, ya que se conocían de antemano los parámetros invariantes y no-invariantes. Sin embargo, este conocimiento no se tiene en aplicaciones reales, donde sería necesario llevar a cabo un proceso iterativo como el propuesto por Asparouhov y Muthén (2014), lo que podría influir en los resultados de potencia al complejizar el procedimiento. En segundo lugar, se han empleado tamaños del efecto pequeños, basados en las diferencias estandarizadas, pero estos no tienen por qué ser los tamaños del efecto más adecuados para realizar todos los contrastes. Puede interesar considerar tamaños del efecto más pequeños si las aplicaciones de los instrumentos tienen consecuencias importantes para los individuos, ya que en estos casos diferencias mínimas podrían ser más relevantes (Borsboom, 2006). Por último, este es el primer trabajo que investiga el funcionamiento de un procedimiento *ad-hoc* para el alineamiento basado en tamaños del efecto. Los resultados obtenidos, por tanto, se reducen a unas condiciones limitadas y sería necesario estudiarlo en un número más amplio de situaciones para comprobar si es una alternativa válida al contraste basado en significación estadística y examinar cómo determinados factores (por ejemplo, el número de grupos y el tamaño muestral) pueden influir en sus resultados.

Referencias

- Asparouhov, T., y Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495-508. doi:10.1080/10705511.2014.919210
- *Barber, C., y Ross, J. (2017). *Arts Participation as a Form of Voluntary Association Membership: A Unique Predictor of Adolescents' Civic Behaviors and Attitudes*. En preparación. Facultad de Educación, Universidad de Missouri.
- *Barber, C., y Ross, J. (2018). Cross-Cohort Changes in Adolescents' Civic Attitudes from 1999 to 2009: An Analysis of Sixteen Countries. *Child Indicators Research*, 11(2), 681-703. <https://doi.org/10.1007/s12187-017-9452-0>
- *Bofah, E. A. T., y Hannula, M. S. (2016). Students' Views on Mathematics in Single-Sex and Coed Classrooms in Ghana. *European Journal of Science and Mathematics Education*, 4(2), 229-250.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, 44(11), S176-S181. doi: 10.1097/01.mlr.0000245143.08679.cc
- *Bratt, C., Abrams, D., Swift, H. J., Vauclair, C. M., y Marques, S. (2018). Perceived age discrimination across age in Europe: From an ageing society to a society for all ages. *Developmental Psychology*, 54(1), 167-180. <http://dx.doi.org/10.1037/dev0000398>
- *Buist, K. L., van Tergouw, M. S., Koot, H. M., y Branje, S. (2019). Longitudinal Linkages between Older and Younger Sibling Depressive Symptoms and Perceived Sibling Relationship Quality. *Journal of Youth and Adolescence*, 48(6), 1190-1202. <https://doi.org/10.1007/s10964-019-01009-y>
- Byrne, B. M. (2016). Adaptation of assessment scales in cross-national research: Issues, guidelines, and caveats. *International Perspectives in Psychology: Research, Practice, Consultation*, 5(1), 51-65. <http://dx.doi.org/10.1037/ipp0000042>
- Byrne, B. M., Shavelson, R. J., y Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456-466.
- Byrne, B. M., y van de Vijver, F. J. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of

- nonequivalence. *International Journal of Testing*, 10(2), 107-132. <https://doi.org/10.1080/15305051003637306>.
- Byrne, B. M., y van de Vijver, F. J. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema*, 29(4), 539-551. doi: 10.7334/psicothema2017.178.
- 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. <http://dx.doi.org/10.1037/a0013193>.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.
- *Coromina, L., y Peral, E. B. (2018). Has the Economic Crisis Eroded Trust in Political Institutions in Europe? A Comparative Analysis from 2008 to 2012. *Visions and Revisions of Europe*, 4, 57-74. doi: 10.17875/gup2018-1115
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., y Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40, 55-75.
- Davidov, E., Muthén, B., y Schmidt, P. (2018). Measurement invariance in cross-national studies: challenging traditional approaches and evaluating new ones. *Sociological Methods & Research*, 47(4), 631–636. <https://doi.org/10.1177/0049124118789708>
- *de Bondt, N., y Van Petegem, P. (2015). Psychometric evaluation of the Overexcitability Questionnaire-Two applying Bayesian Structural Equation Modeling (BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Frontiers in Psychology*, 6, 1963. <https://doi.org/10.3389/fpsyg.2015.01963>
- *de Carvalho, J. S., Pereira, N. S., Pinto, A. M., y Marôco, J. (2016). Psychometric properties of the mental health continuum-short form: a study of Portuguese speaking children/youths. *Journal of Child and Family Studies*, 25(7), 2141-2154. <https://doi.org/10.1007/s10826-016-0396-7>
- Desa, D., van de Vijver, F. J. R., Carstens. R., y Schulz, W. (2019). Measurement invariance in international large-scale assessments: Integrating theory and method. En T. P. Johnson, B. E. Pennell, I. Stoop, y B. Dorer (Eds.), *Advances in comparative survey methodology* (pp. 881-910). New York, NY: Wiley

- *Duell, N., Steinberg, L., Chein, J., Al-Hassan, S. M., Bacchini, D., Lei, C., ... y Lansford, J. E. (2016). Interaction of reward seeking and self-regulation in the prediction of risk taking: A cross-national test of the dual systems model. *Developmental Psychology*, *52*(10), 1593 -1605. <http://dx.doi.org/10.1037/dev0000152>
- *Elsworth, G. R., Beauchamp, A., y Osborne, R. H. (2016). Measuring health literacy in community agencies: a Bayesian study of the factor structure and measurement invariance of the health literacy questionnaire (HLQ). *BMC Health Services Research*, *16*(1), 508. <https://doi.org/10.1186/s12913-016-1754-2>
- *Evers, A., McCormick, C. M., Hawley, L. R., Muñoz, J., Balboni, G., Bartram, D., ... y Fine, S. (2017). Testing practices and attitudes toward tests and testing: An international survey. *International Journal of Testing*, *17*(2), 158-190. <https://doi.org/10.1080/15305058.2016.1216434>
- Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532-538. <http://dx.doi.org/10.1037/a0015808>
- Finch, W. H. (2016). Detection of differential item functioning for more than two groups: A Monte Carlo comparison of methods. *Applied Measurement in Education*, *29*(1), 30-45. <https://doi.org/10.1080/08957347.2015.1102916>
- Flake, J. K., y 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>
- *Gnambs, T., y Appel, M. (2019). Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. *Computers in Human Behavior*, *93*, 53-61. <https://doi.org/10.1016/j.chb.2018.11.045>
- *Gelissen, J. (2018). Testing measurement equivalence of experienced holiday quality: Evidence on built-in bias in the Flash Eurobarometer survey data. *Tourism Management*, *67*, 273-283. <https://doi.org/10.1016/j.tourman.2018.02.005>
- *Goel, A., y Gross, A. (2019). Differential item functioning in the cognitive screener used in the Longitudinal Aging Study in India. *International Psychogeriatrics*, 1-11. <https://doi.org/10.1017/S1041610218001746>

- *Goroshit, M. (2016). Political Participation: A Latent Variable Approach. Testing Measurement Equivalence of Political Participation Using ESS Data. *Eurasian Journal of Social Sciences*, 4(1), 26-38. doi: 10.15604/ejss.2016.04.01.003
- *Guio, A. C., Gordon, D., Najera, H., y Pomati, M. (2017). Revising the EU material deprivation variables. Informe técnico, EUROSTAT. Disponible en: <https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/KS-TC-17-002?inheritRedirect=true>
- Horn, J. L., y McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research: An International Journal Devoted to the Scientific Study of the Aging Process*, 18(3), 117-144. doi: 10.1080/03610739208253916
- *Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Guo, J., Althoff, R. R., Kan, K. J., ... y da Rocha, M. M. (2019). Testing syndromes of psychopathology in parent and youth ratings across societies. *Journal of Clinical Child & Adolescent Psychology*, 48(4), 596-609. <https://doi.org/10.1080/15374416.2017.1405352>
- *Jang, S. (2015). A Multilevel Examination of Cultural Moderators of the Job Demands-Resources Model. *Graduate Theses and Dissertations*. <https://scholarcommons.usf.edu/etd/5826>
- *Jang, S., Kim, E. S., Cao, C., Allen, T. D., Cooper, C. L., Lapierre, L. M., ... y Abarca, N. (2017). Measurement invariance of the satisfaction with life scale across 26 countries. *Journal of Cross-Cultural Psychology*, 48(4), 560-576. <https://doi.org/10.1177/0022022117697844>
- *Jayawickreme, N., Verkuilen, J., Jayawickreme, E., Acosta, K., y Foa, E. B. (2017). Measuring depression in a non-Western war-affected displaced population: measurement equivalence of the beck depression inventory. *Frontiers in Psychology*, 8, 1670. <https://doi.org/10.3389/fpsyg.2017.01670>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., y Rosseel, Y. (2018). semTools: Useful tools for structural equation modeling. R package version 0.5-0. Disponible en: <https://CRAN.R-project.org/package=semTools>
- Kim, E. S., Cao, C., Wang, Y., y Nguyen, D. T. (2017). Measurement invariance testing with many groups: a comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 524-544. DOI: 10.1080/10705511.2017.1304822

- Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Nueva York: Guilford.
- *Kotzur, P. F., Friehs, M. T., Asbrock, F., y van Zalk, M. H. (2019). Stereotype content of refugee subgroups in Germany. *European Journal of Social Psychology*, 1–15. <https://doi.org/10.1002/ejsp.2585>
- *Kwiatkowska, M. M., Rogoza, R., y Volkodav, T. (2018). Psychometric properties of the Benign and Malicious Envy Scale: Assessment of structure, reliability, and measurement invariance across the United States, Germany, Russia, and Poland. <https://doi.org/10.31234/osf.io/7wutz>
- *Lamm, R., Do, T., y Rodriguez, M. C. (2019, abril). *Measurement Invariance of an International Developmental Assets Measure: Alignment of 29 Countries*. Artículo presentado en la reunión anual del Consejo Nacional de Medición en Educación, Toronto, Canadá.
- Lek K, Oberski D, Davidov E, Cieciuch J, Seddig D, Schmidt P. (2019). Approximate Measurement Invariance. En Johnson T. P., Pennell B. E., Stoop I. A. L., Dorer B., (Eds.) *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, (pp. 911-928). Hoboken: John Wiley & Sons, Inc.
- *Li, J. B., Delvecchio, E., Lis, A., y Mazzeschi, C. (2018). Family allocentrism and its relation to adjustment among Chinese and Italian adolescents. *Psychiatry Research*, 270, 954-960. <https://doi.org/10.1016/j.psychres.2018.03.036>
- *Li, J. B., Salcuni, S., y Delvecchio, E. (2019). Meaning in life, self-control and psychological distress among adolescents: A cross-national study. *Psychiatry research*, 272, 122-129. <https://doi.org/10.1016/j.psychres.2018.12.033>
- *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>
- *Lomazzi, V., Israel, S., y Crespi, I. (2019). Gender Equality in Europe and the Effect of Work-Family Balance Policies on Gender-Role Attitudes. *Social Sciences*, 8(1), 1-29. <https://doi.org/10.3390/socsci8010005>

- *Lomos, C. (2017). To what extent do teachers in European countries differ in their professional community practices? *School Effectiveness and School Improvement*, 28(2), 276-291. <http://dx.doi.org/10.1080/09243453.2017.1279186>
- MacCallum, R. C., Roznowski, M., y Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin*, 111(3), 490-504.
- *Mansur, J., Sobral, F., y Goldszmidt, R. (2017). Shades of paternalistic leadership across cultures. *Journal of World Business*, 52(5), 702-713. <https://doi.org/10.1016/j.jwb.2017.06.003>
- Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., y Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, 23(3), 524-545. <http://dx.doi.org/10.1037/met0000113>.
- *McGrath, R. E. (2015). Measurement invariance in translations of the VIA inventory of strengths. *European Journal of Psychological Assessment*, 32(3), 187–194. <https://doi.org/10.1027/1015-5759/a000248>
- Meade, A. W., y Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 611-635. <https://doi.org/10.1080/10705510701575461>
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127-143. [https://doi.org/10.1016/0883-0355\(89\)90002-5](https://doi.org/10.1016/0883-0355(89)90002-5)
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543. <https://doi.org/10.1007/BF02294825>
- Meredith, W., y Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, 44(11), S69-S77. Disponible en: <http://www.jstor.org/stable/41219507>
- *Milfont, T. L., Bain, P. G., Kashima, Y., Corral-Verdugo, V., Pasquali, C., Johansson, L. O., ... y Bilewicz, M. (2018) On the relation between social dominance

- orientation and environmentalism: A 25-nation study. *Social Psychological and Personality Science*, 9(7), 802-814. <https://doi.org/10.1177/1948550617722832>
- Millsap, R. E., y Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, 9(1), 93-115. doi: 10.1037/1082-989X.9.1.93
- Millsap, R. E., y Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. En *Factor analysis at 100. Historical Developments and Future Directions*, eds. R Cudeck, RC MacCallum, 8: 131-152. Mahwah (NJ): Lawrence Erlbaum.
- *Munck, I., Barber, C., y 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. <https://doi.org/10.1177/0049124117729691>
- Muthén, B. O. (2002). Using Mplus Monte Carlo simulations in practice: A note on assessing estimation quality and power in latent variable models. Mplus web notes, No. 1 Version 2. Disponible en: <http://www.statmodel.com/download/webnotes/mc1.pdf>.
- Muthén, B., y Asparouhov, T. (2013). New methods for the study of measurement invariance with many groups. Informe técnico. Disponible en: <http://statmodel2.com/download/PolAn.pdf>.
- Muthén, B., y Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology*, 5, 978. <https://doi.org/10.3389/fpsyg.2014.00978>
- Muthén, L. K. y Muthén, B. O. (1998-2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén
- *Najera, H. E. (2017). Does measurement invariance hold for the official Mexican multidimensional poverty measure? A state-level analysis 2012. *Quality & Quantity*, 51(3), 1217-1241. <https://doi.org/10.1007/s11135-016-0327-0>
- *Phelan, S., Bauer, J., y Lewalter, D. (2018). Visit motivations: development of a short scale for comparison across sites. *Museum Management and Curatorship*, 33(1), 25-41. <https://doi.org/10.1080/09647775.2017.1389617>

- *Pokropek, A., Borgonovi, F., y Jakubowski, M. (2015). Socio-economic disparities in academic achievement: A comparative analysis of mechanisms and pathways. *Learning and Individual Differences*, 42, 10-18. <https://doi.org/10.1016/j.lindif.2015.07.011>
- Pokropek, A., Davidov, E., y Schmidt, P. (2019). A Monte Carlo Simulation Study to Assess The Appropriateness of Traditional and Newer Approaches to Test for Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-21. <https://doi.org/10.1080/10705511.2018.1561293>
- Pornprasertmanit, S. (2018). A Note on Effect Size for Measurement Invariance. Disponible en: <http://cran.irsn.fr/web/packages/semTools/vignettes/partialInvariance.pdf>
- Reinartz, W., Haenlein, M., y Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of research in Marketing*, 26(4), 332-344. <https://doi.org/10.1016/j.ijresmar.2009.08.001>
- Robitzsch, A. (2019). sirt: Supplementary Item Response Theory Models. R package version 3.6-21. Disponible en: <https://CRAN.R-project.org/package=sirt>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. Disponible en: <http://www.jstatsoft.org/v48/i02/>.
- *Russell, J. D., Graham, R. A., Neill, E. L., y Weems, C. F. (2016). Agreement in youth-parent perceptions of parenting behaviors: A case for testing measurement invariance in reporter discrepancy research. *Journal of Youth and Adolescence*, 45(10), 2094-2107. <https://doi.org/10.1007/s10964-016-0495-1>
- Rutkowski, L., y 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>.
- *Silva, B. C. (2017). *Contemporary populism: Actors, causes, and consequences across 28 democracies*. Tesis doctoral, Universidad Centroeuropea.
- *Silva, B. C., Andreadis, I., Anduiza, E., Blanuša, N., Corti, Y. M., Delfino, G., ... y Littvay, L. (2018). Public opinion surveys: A new scale. En Hawkins, K. A.,

Carlin, R. E., Littvay, L., y Kaltwasser, C. R. (Eds.), *The Ideational Approach to Populism* (pp. 150-177). Routledge.

- *Smale, A., Bagdadli, S., Cotton, R., Dello Russo, S., Dickmann, M., Dysvik, A., Gianecchini, M., Kaše, R., Lazarova, M., Reichel, A., y Rozo, P. (2019). Proactive career behaviors and subjective career success: The moderating role of national culture. *Journal of Organizational Behavior*, 40(1), 105-122. <https://doi.org/10.1002/job.2316>
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27(2), 229-239. <https://doi.org/10.1111/j.2044-8317.1974.tb00543.x>
- Steenkamp, J. B. E., y Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78-90. <https://doi.org/10.1086/209528>.
- Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology*, 9(1), 1-12. doi: 10.1027/1614-2241/a000049
- *Su, H. C., Chen, Y. S., y Kao, T. W. (2018). Enhancing Supplier Development: An Efficiency Perspective. *Journal of Business Logistics*, 39(4), 248-266. <https://doi.org/10.1111/jbl.12197>
- *Tay, A. K., Jayasuriya, R., Jayasuriya, D., y Silove, D. (2017). Measurement invariance of the Hopkins Symptoms Checklist: a novel multigroup alignment analytic approach to a large epidemiological sample across eight conflict-affected districts from a nation-wide survey in Sri Lanka. *Conflict and health*, 11(1), 8. <https://doi.org/10.1186/s13031-017-0109-x>
- van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., y Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in Psychology*, 6, 1064. <https://doi.org/10.3389/fpsyg.2015.01064>
- *van Holland De Graaf, J., Hoogenboom, M., De Roos, S., y Bucx, F. (2018). Sociodemographic correlates of fathers' and mothers' parenting behaviors. *Journal of Child and Family Studies*, 27(7), 2315-2327. <https://doi.org/10.1007/s10826-018-1059-7>

- *Vuyk, M. A., Krieshok, T. S., y Kerr, B. A. (2016). Openness to experience rather than overexcitabilities: Call it like it is. *Gifted Child Quarterly*, 60(3), 192-211. <https://doi.org/10.1177/0016986216645407>
- *Wei, J., Pomerantz, E. M., Ng, F. F. Y., Yu, Y., Wang, M., y Wang, Q. (2019). Why does parents' involvement in youth's learning vary across elementary, middle, and high school?. *Contemporary Educational Psychology*, 56, 262-274. <https://doi.org/10.1016/j.cedpsych.2018.12.007>
- *Weziak-Bialowolska, D. (2015). Differences in gender norms between countries: Are they valid? The issue of measurement invariance. *European Journal of Population*, 31(1), 51-76. doi: 10.1007/s10680-014-9329-6
- *Woitschach, P. (2018). *Evaluaciones educativas a gran escala en Latinoamérica: TERCE*. Tesis doctoral, Facultad de Psicología, Universidad Complutense de Madrid.
- *Yu, S., Zhang, F., Nunes, L. D., y Levesque-Bristol, C. (2018). Self-determined motivation to choose college majors, its antecedents, and outcomes: A cross-cultural investigation. *Journal of Vocational Behavior*, 108, 132-150. doi:10.1016/j.jvb.2018.07.002
- *Yuen, C. X. (2016). *Adolescents' and mothers' affect during their daily interactions: an examination in the United States and China*. Tesis doctoral, Universidad de Illinois en Urbana-Champaign.
- *Żemojtel-Piotrowska, M., Piotrowski, J. P., Ciecuch, J., Adams, B. G., Osin, E. N., Ardi, R., et al. (2017). Measurement invariance of personal well-being index (PWI-8) across 26 countries. *Journal of Happiness Studies*, 18(6), 1679–1711. doi: 10.1007/s10902-016-9795-0
- *Zickfeld, J. H., Schubert, T. W., Seibt, B., Blomster, J. K., Arriaga, P., Basabe, N., ... y Ding, Y. (2018). Kama muta: Conceptualizing and measuring the experience often labelled being moved across 19 nations and 15 languages. *Emotion*, 19(3), 402-424. <http://dx.doi.org/10.1037/emo0000450>

Apéndice A

Tabla resumen de las características de los estudios que han empleado el método de alineamiento

| Tabla A1 Características de los estudios que han aplicado el método de alineamiento | | | | | | | |
|--|----------------------------------|--|---|---------------------------------|---|----------------------|----------------------------------|
| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
| Bofah y Hannula (2016) | 4 (2 tipos de escuela x 2 sexos) | 2034 estudiantes (n _g = 394 - 783). | <i>View of Mathematics (VoM)</i> : 25 ítems y 4 factores (con 12, 6, 4 y 3 ítems cada uno) | AFC-MG | Fijo | - | Alineamiento |
| Bratt, Abrams, Swift, Vanclair y Marques (2018) | 29 (países) | 56272 personas de 15 a 105 años | 3 ítems (1 factor) de discriminación por edad de la Encuesta Social Europea (ESS) | Alineamiento, LSEM y MNFLA | Fijo | ML | MNFLA y alineamiento |
| Yu, Zhang, Nunes y Levesque-Bristol (2018) | 2 (China y EEUU) | 146 estudiantes de EEUU y 479 estudiantes emparejados de China | 1) IAF: 9 ítems y 3 factores (3 ítems cada uno); 2) POPS-M: 4 ítems (1 factor); 3) elección autodeterminada: 12 ítems y 2 factores (6 ítems cada uno) y 4) motivación académica autodeterminada: 12 ítems y 2F (6 ítems cada uno) | AFC-MG | Fijo | - | Alineamiento |
| Zickfel et al. (2018) | 19 (países) | 3542 adultos (n _g = 45-332) | <i>Kama Muta Multiplex Scale (KAMMUS)</i> : 50 ítems y 5 factores. Solo se evalúan las escalas de appraisal (11 ítems), motivación (7 ítems) y etiquetas emocionales (7 ítems) | Alineamiento | Libre (appraisal) y fijo (motivación y etiquetas) | MLR | - |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|------------------------------|---|--|---|---------------------------------|----------------------|----------------------|----------------------------------|
| Weziak-Bralowolska (2015) | 16 (países) | Datos procedentes de la WVS (n _g = 650-1239) | <i>Gender Equality Scale (GES)</i> : 5 ítems (1 factor) | AFC-MG y alineamiento | - | MLR | AFC-MG y alineamiento |
| Vuyk, Krieshok y Kerr (2016) | 2 (talentosos vs. población general) | 149 adultos y adolescentes con talento creativo e intelectual y 312 adultos de población general | 1) Factor O de NEO-PI-3: 6 factores con 10 ítems cada uno); 2) <i>Overexcitability Questionnaire-Two</i> (OEQ-II): 50 ítems y 5F (con 10 ítems cada uno) | AFC-MG y alineamiento | Fijo | - | - |
| Smale et al. (2019) | 22 (países) | 11892 adultos trabajadores | 1) Éxito financiero: 1 factor con 3 ítems; 2) balance vida-trabajo: 1 factor con 3 ítems; 3) EMA: 1 factor con 5 ítems | Alineamiento | - | - | - |
| Barber y Ross (2018) | 66 (2 género x 16 país x 2 cohorte + 2 género CIVED EEUU) | 116155 adolescentes; datos procedentes del CIVED (1999) y del ICCS (2009) (n _g = 2544-6192) | 5 escalas actitudinales (1 factor cada una, con 5, 6, 4, 4 y 6 ítems cada uno) del CIVED y el ICCS (solo se incluyen los ítems comunes a ambas encuestas) | Alineamiento | Fijo | MLR | Regresión múltiple |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|---|---------------------------------|---|---|--------------------------------------|----------------------|----------------------|--|
| Russell, Graham, Neill y Weems (2016) | 4 (hijos x padres) | 125 hijos niños, 130 hijos adolescentes, 125 padres de niños y 130 padres de adolescentes | <i>Alabama Parenting Questionnaire</i> (APQ): 42 ítems y 5 factores | AFC-MG y alineamiento | - | - | Prueba de diferencias de medias latentes |
| Milfont et al. (2018) | 25 (países) y 10 (países) | 4163 estudiantes (de 25 países) y 1237 participantes de población general (de 10 países) ($n_g = 38-267$) | Escala corta de orientación a la dominancia social (SSDO): 4 ítems (1 factor) | Rotación procrusteana y alineamiento | Fijo | ML | Alineamiento (suplementario) |
| Munck, Barber y Tomey-Purta (2018) | 92 ((22+24 países) x 2 géneros) | 79278 adolescentes; datos procedentes del CIVED99 y del ICCS09 ($n_g = 226-1689$) | <i>Support for Immigrants' Rights</i> (SIR): 5 ítems (1 factor) (ítems iguales en CIVED99 e ICCS09) | AFC-MG y alineamiento | Fijo | ML | Alineamiento |
| Pokropek, Borgonovi y Jakubowski (2015) | 33 (países) | 290361 estudiantes de 15-16 años ($n_g = 3508-33806$) | Medida de recursos domésticos: 13 ítems y 3 factores (con 5, 4 y 4 ítems cada uno) | AFC-MG y alineamiento | - | - | - |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|--|----------------------|--|--|---|----------------------------|--|----------------------------------|
| Li, Delvecchio, Lis y Mazzeschi (2018) | 2 (China e Italia) | 484 adolescentes chinos y 480 adolescentes italianos | <i>Family Allocentrism Scale (FAS)</i> : 21 ítems (1 factor) | AFC-MG y alineamiento | Fijo y libre | MLR | Alineamiento |
| Jayawickreme, Verkuilen, Jayawickreme, Acosta y Foa (2017) | 2 (Sri Lanka y EEUU) | 268 mujeres de EEUU y 149 mujeres de Sri Lanka | <i>Beck Depression Inventory (BDI)</i> : 21 ítems y 3 factores (11, 2 y 3 ítems) ó 2 factores (13 y 8 ítems) | Alineamiento y <i>bifactor</i> exploratorio | - | MLR | - |
| de Bondt y van Petegem (2015) | 2 (géneros) | 318 mujeres jóvenes y 198 hombres jóvenes | <i>Overexcitability Questionnaire-Two (OEQ-II)</i> : 50 ítems y 5F (con 10 ítems cada uno) | Alineamiento | - | Bayesiano (varianza previa $\leq 0,01$) | Alineamiento |
| Duell et al. (2016) | 11 (países) | 5227 participantes ($n_g = 364-556$) | 1) Toma de riesgos (RT): 9 ítems y 1 factor; 2) búsqueda de sensaciones (SS): 6 ítems y 1 factor; y 3) planificación (P): 6 ítems y 1 factor | Alineamiento | Fijo (RT y P) y libre (SS) | ML | - |
| Jang et al. (2017) | 26 (países) | 7004 managers ($n_g = 137-500$) | <i>Satisfaction With Life Scale (SWLS)</i> : 5 ítems (1 factor) | AFC-MG, AFC-ML y alineamiento | Fijo | ML | AFC-MG y alineamiento |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|---|--|---|--|---------------------------------|----------------------|----------------------|---|
| Elsworth, Beauchamp y Osborne (2016) | 8 (agencias), 2 (género), 2 (lenguaje), 5 (quintiles de edad) y 5 (categorías nivel educativo) | 813 clientes de 8 agencias comunitarias (n _g = 70-120) | <i>Health Literacy Questionnaire</i> (HLQ): 44 ítems y 9 factores (con entre 4 y 6 ítems cada uno) | AFC-MG y alineamiento | Libre | Bayesiano | - |
| Lomos (2017) | 22 (países) | 35897 profesores (n _g = 112-2964) | 5 ítems de comunidad profesional del ICCS (1 factor) | Alineamiento | Fijo | ML | Diversos métodos, entre ellos, alineamiento |
| van Holland de Graaf, Hoogenboom, De Ross y Buck (2018) | 2 (género) | 903 padres y 1197 madres de niños y adolescentes | 6 escalas de estrategias parentales: 1) afecto (5 ítems); 2) responsividad (6 ítems); 3) explicación (5 ítems); 4) apoyo a la autonomía (6 ítems); 5) recompensa (5 ítems) y 6) castigo (6 ítems) | Alineamiento | - | - | - |
| Lomazzi (2018) | 59 (países) | 89320 participantes (n _g = 841-5659) | 5 ítems sobre actitudes sobre roles de género de la Encuesta Mundial de Valores (WVS) (1 factor) | AFC-MG y alineamiento | Fijo | ML | Alineamiento |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|----------------------|--|--|--|---------------------------------|----------------------|----------------------|----------------------------------|
| Goroshit (2016) | 26 (países) | 50781 participantes (n _g = 1083-3031) | 6 ítems de participación política de la ESS (2 factores con 3 ítems cada uno) | AFC-MG y alineamiento | Fijo | - | - |
| Barber y Ross (2017) | 28 (países CIVED99) y 38 (países ICCS09) | 52654 estudiantes del CIVED99, y 119847 estudiantes del ICCS09 (n _g = 357-6576) | 1) 14 ítems de participación extracurricular del CIVED99; y 2) 11 ítems de participación extracurricular del ICCS09 | Alineamiento | Fijo | MLR | - |
| McGrath (2015) | 16 (países) | 15540 participantes (n _g = 714-1000) | VIA <i>Inventory of Strengths</i> (VIA-IS): 240 ítems y 5 factores (cada factor tiene 6, 5, 5, 5 y 3 facetas, con 10 ítems cada faceta) | AFC-MG y alineamiento | Fijo | - | - |
| Yuen (2016) | 2 (China y EEUU) | 1) 80 madres y 80 hijos de EEUU; y 2) 71 madres y 71 hijos de China | 1) Entrevista de afecto interactivo: 6 ítems y 2 factores (con 3 ítems cada uno); 2) entrevista de afecto general: 8 ítems y 2 factores (con 4 ítems cada uno); 3) encuesta de afecto general: 9 y 11 ítems y 2 factores (ítems agrupados en 3 parcelas) | Alineamiento | Fijo | ML | Modelos mixtos |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|-----------------------|--|---|--|---------------------------------|----------------------|----------------------|----------------------------------|
| Ivanova et al. (2018) | 30 (sociedades CBL/6-18) y 19 (sociedades YSR) | 61703 participantes para el CBL/6-18; 29489 adolescentes para el YSR (n _g = 946-4848) | 1) <i>Child Behavior Checklist</i> (CBL/6-18): 96 ítems y 8 factores (con 8-18 ítems por factor); y 2) <i>Youth Self-Report</i> (YSR): 88 ítems y 8 factores 8 factores: 1) restricciones organizacionales (5 ítems); 2) control (4 ítems); 3) participación en decisiones (3 ítems); 4) apoyo del supervisor (6 ítems); 5) apoyo del líder (4 ítems); 6) metas y feedback (3 ítems); 7) satisfacción (4 ítems); y 8) intención de rotación (2 ítems) | Alineamiento | - | Bayesiano robusto | - |
| Jang (2015) | 28 (países) | 24385 trabajadores (n _g = 231-1028) | | Alineamiento | Fijo | ML | - |
| Silva (2017) | 4 (2 tipos de apoyo x 2 años) | Datos procedentes del Latinobarómetro (2015) sobre apoyo institucional en dos momentos distintos: antes (2005) y después (2006) de las elecciones (n _g > 1000) | Medida de apoyo institucional: 6 indicadores y 1 factor | AFC-MG y alineamiento | - | - | - |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|---------------------------------|---------------------------|---|---|---------------------------------|----------------------|----------------------|----------------------------------|
| Silva et al. (2018) | 8 (países) | 1681 participantes para las fases 1 y 2 del estudio (n _g = 153-262) | Fase 2 del desarrollo de una escala de populismo: 4 factores con 12 ítems cada uno (los que más pesan en su factor) | Alineamiento | - | - | - |
| Woitschach (2018) | 16 (15 países + 1 estado) | 61921 estudiantes (n _g = 406-846) | 82 ítems de conocimiento científico del TERCE divididos en 6 cuadernillos (con un factor y 26-30 ítems por cada cuadernillo) | Alineamiento | Libre y fijo | MLR | Alineamiento |
| Li, Saleuni y Delvecchio (2019) | 2 (China e Italia) | 270 adolescentes de China y 219 adolescentes de Italia | <i>Meaning in Life Questionnaire</i> (MLQ): 10 ítems y 2 factores (5 ítems cada uno) | AFC-MG y alineamiento | Fijo | - | Alineamiento |
| Su, Chen y Kao (2018) | 11 (países) | Datos procedentes del proyecto HPM de 261 plantas de manufacturación (n _g ≥ 100) | 1) 4 ítems de desarrollo del proveedor; 2) 5 ítems de evaluación del proveedor; 3) 10 ítems de rendimiento del proveedor; 4) 3 ítems de normas relacionales; 5) 7 ítems de compartir información por parte del comprador; y 6) 7 ítems de compartir información por parte del proveedor | Alineamiento | - | - | - |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|---|--|---|--|---------------------------------|----------------------|----------------------|----------------------------------|
| Coromina y Peral (2018) | 23 (3 períodos x 8 países) | Datos procedentes de la ESS (n _g = 1544-3015) | 3 ítems de la ESS sobre confianza en las instituciones (1 factor) | AFC-MG y alineamiento | - | - | AFC-MG y alineamiento |
| Goel y Gross (2019) | 2 (nivel de alfabetización), 2 (nivel educativo) y 2 (rural vs. urbano) | 739 participantes de 60 años o más (n _g = 202-537) | Screening cognitivo breve del Estudio Longitudinal de Envejecimiento en India (LASI): 11 ítems (1 factor) | Alineamiento | - | ML | - |
| Lamm, Do, Rodriguez, Scales y Roehlkertain (2019) | 29 (países) | 24317 jóvenes (n _g = 100 - 6926) | 9 ítems (1 factor) de sensación de pertenencia de la escala de Recursos para el Desarrollo (DAP) | Alineamiento | Fijo | - | Alineamiento |
| Buist, van Tergouw, Koot y Branje (2019) | 6 (recogidas) | 367 díadas de hermanos | 1) Subescalas de estado disfórico (8 ítems), autoevaluación negativa (8 ítems) y quejas somáticas (7 ítems) de la Escala de Depresión Adolescente de Reynolds (RADS-II); y 2) subescalas de apoyo (12 ítems), interacción negativa (6 ítems) y poder (6 ítems) del Inventario de Red de Relaciones (NRI) | Alineamiento | - | - | - |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|---|--|---|---|--|----------------------|----------------------|------------------------------------|
| Gnambis y Appel (2019) | 27 (países) y 3 (recogidas) | 80396 participantes (n _g = 500 - 1572) | Medida de evaluación general de los robots de 3 ítems (1 factor) | AFC-MG (entre recogidas) y alineamiento (entre países) | - | - | Tamaño del efecto y regresiones |
| de Carvalho, Pereira, Pinto y Marôco (2016) | 2 (nivel de enseñanza), 3 (grupo de edad) y 2 (género) | 208 estudiantes de enseñanza elemental (7-9 años) y 216 estudiantes de enseñanza media (10-14 años) | <i>Mental Health Continuum-Short Form</i> (MHC-SF): 14 ítems y 3 factores (con 3, 5 y 6 ítems cada uno) | AFC-MG y alineamiento | - | - | Pruebas T |
| Phelan, Baver y Lewalter (2018) | 3 (contextos de ocio educativo) | 99 asistentes a museo de ciencias, 190 asistentes a museo de artes y 216 asistentes a zoo | Desarrollan un test de 17 ítems y 6 factores (con entre 2 y 4 ítems cada uno) | AFC-MG y alineamiento | Fijo | - | Estadístico de Wald y alineamiento |
| Mansur, Sobral y Golszmith (2017) | 10 (clusters de países) | 13646 participantes (n _g = 759-1758) | Medida de liderazgo paternalista derivada del proyecto GLOBE: 11 ítems y 3 factores (con 4, 3 y 4 ítems cada uno) | AFC-MG y alineamiento | - | ML | Diversas técnicas |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|---|---|--|---|---------------------------------|----------------------|----------------------|----------------------------------|
| Gelissen (2018) | 35 (países 2014), 33 (países 2015 y 2016) y 3 (recogidas) | 28435 participantes (n _g = 114-1845) | Medida de calidad de vacaciones experimentada (EHQ): 7 ítems (1 factor) (ítems iguales en las tres recogidas) | AFC-MG y alineamiento | - | - | Alineamiento |
| Tay, Jayasuriya, Jayasuriya y Silove (2017) | 8 (distritos) | 8456 encuestados (n _g = 1016-1112) | <i>Hopkins Symptoms Checklist</i> (HSCL-25): 25 ítems y 2 factores | Alineamiento | Fijo | - | Alineamiento |
| Lomazzi, Israel y Crespi (2019) | 28 (países) | 16843 encuestados (n _g = 304-1043) | 4 ítems (1 factor) sobre actitudes de género 1) Involucración de los padres en el aprendizaje: 2 factores con 5 y 8 ítems cada uno; 2) invitaciones de los hijos a la involucración: 2 factores con 5 y 8 ítems cada uno; 3) creencias de los padres sobre el rol de implicación: 1 factor con 6 ítems; 4) creencias de los padres sobre autoeficacia: 1 factor con 9 ítems; y 5) normas percibidas sobre la involucración: 2 factores con 3 y 4 ítems cada uno | AFC-MG y alineamiento | - | - | - |
| Wei et al. (2019) | 3 (etapas escolares) | 3074 jóvenes y 3074 padres (n _g = 830-1224) | | AFC-MG y alineamiento | - | - | Modelos SEM |

Tabla A1 (continuación)
Características de los estudios que han aplicado el método de alineamiento

| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
|---|------------------------|--|---|---------------------------------|----------------------|----------------------|----------------------------------|
| Kwiatkowska, Rogoza y Volkodav (2018) | 4 (países) | 2792 residentes (n _g = 558-799) | <i>Benign and Malicious Envy Scale</i> (BeMas): 10 ítems y 2 factores (5 ítems cada uno) | AFC-MG y alineamiento | - | - | - |
| Kotzur, Friehs, Asbrock y van Zalk (2019) | 10 (grupos sociales) | 246 adultos universitarios alemanes | Medida del Modelo de Contenido de Estereotipos (SCM) de 6 ítems (2 factores con 3 ítems cada uno) | AFC-MG y alineamiento | Fijo | MLR | Alineamiento |
| Žemojtel-Piotrowska et al. (2017) | 26 (países) | 5530 estudiantes universitarios (n _g = 161-302) | <i>Personal Well-being Index</i> (PWI-8): 8 ítems (1 factor) <i>EFPA Questionnaire on Test Attitudes of Psychologists</i> (EQTAP): 29 ítems y 5 factores (con 8, 8, 6, 4 y 3 ítems cada uno) | AFC-MG y alineamiento | - | - | - |
| Evers et al. (2016) | 29 (países) | 20467 psicólogos (n _g = 20-5482) | | Alineamiento | Fijo | MLR | Modelos multinivel |
| Najera (2017) | 32 (estados de México) | 212674 encuestados | Dimensión de privación de la <i>Mexican multidimensional poverty measure</i> (MPM): 10 ítems (1 factor). | Alineamiento | Fijo | ML | Alineamiento |
| Guio, Gordon, Najera y Pomati (2017) | 27 (países) | - | Medida de privación material (MD): 13 ítems (1 factor) | AFC-MG y alineamiento | - | - | Alineamiento |

Tabla A1 (continuación)

| Características de los estudios que han aplicado el método de alineamiento | | | | | | | |
|--|------------------------------------|--|--|---|----------------------|----------------------|----------------------------------|
| Estudio | Número de grupos | Muestra total y submuestras grupales | Medidas | Técnica de invarianza utilizada | Tipo de alineamiento | Método de estimación | Técnica de comparación de medias |
| Byrne y van de Vijver (2017) | 27 (países) | 5482 estudiantes universitarios (n _g = 65-450) | <i>Family Values Scale</i> (FVS): 14 ítems y 2 factores (con 6 y 8 ítems) | AFC-MG y alineamiento | Fijo | MLR | Alineamiento |
| Muthén y Asparouhov (2014) | 28 (14 países x 2 cohortes) | 29449 estudiantes del CIVED99 y 10643 estudiantes del ICCS09 | 17 ítems de conocimiento del CIVED99 (1 factor) y 17 ítems de conocimiento del ICCS (1 factor) (ítems comparables entre ambos) | Alineamiento | Fijo | ML | Alineamiento |
| Muthén y Asparouhov (2013) | 26 (países) | 49894 participantes (n _g media = 1919) | 4 ítems de tradición-conformidad de la ESS (1 factor) | AFC-MG, alineamiento y AFC multinivel | Fijo | ML | Alineamiento y AFC multinivel |
| Marsh et al. (2018) | 30 (países) y 60 (países x género) | 249840 estudiantes | Constructos motivacionales del PISA: 44 ítems y 8 factores (con 4-8 ítems cada uno) | Alineamiento y <i>alignment-within-CFA</i> (AwC) | Fijo | MLR | Alineamiento y AwC |
| Kim, Cao, Wang y Nguyen (2017) (suplementario) | 65 (países) | 17303 directores de escuelas (n _g = 11-1458) | 7 ítems de efectividad del personal del PISA (1 factor) | AFC-MG, AFC multinivel, FMM multinivel, BSEM y alineamiento | Libre y fijo | ML | - |

Apéndice B

Tablas resumen de RMSE promedio global para los distintos parámetros

Tabla A2
 RMSE promedio de intersecciones para no-invarianza en pesos

| Método | Número de grupos | Tamaño muestral | $\Delta\lambda = 0$ | $\Delta\lambda = .1$ | $\Delta\lambda = .2$ | $\Delta\lambda = .3$ | $\Delta\lambda = .4$ |
|--------------|------------------|-----------------|---------------------|----------------------|----------------------|----------------------|----------------------|
| AFC-MG | 4 | 250 | .056 | .056 | .056 | .056 | .056 |
| | | 500 | .040 | .040 | .040 | .040 | .040 |
| | | 1000 | .028 | .028 | .028 | .028 | .028 |
| | 8 | 250 | .055 | .055 | .055 | .055 | .055 |
| | | 500 | .039 | .039 | .039 | .039 | .039 |
| | | 1000 | .027 | .027 | .027 | .027 | .027 |
| | 16 | 250 | .055 | .055 | .055 | .055 | .055 |
| | | 500 | .039 | .039 | .039 | .039 | .039 |
| | | 1000 | .027 | .027 | .027 | .027 | .027 |
| | 32 | 250 | .053 | .053 | .053 | .053 | .053 |
| | | 500 | .039 | .039 | .039 | .039 | .039 |
| | | 1000 | .026 | .026 | .026 | .026 | .026 |
| Alineamiento | 4 | 250 | .063 | .063 | .063 | .063 | .063 |
| | | 500 | .044 | .044 | .044 | .044 | .044 |
| | | 1000 | .031 | .031 | .031 | .031 | .030 |
| | 8 | 250 | .063 | .063 | .063 | .063 | .062 |
| | | 500 | .045 | .045 | .045 | .044 | .044 |
| | | 1000 | .031 | .031 | .031 | .031 | .031 |
| | 16 | 250 | .062 | .062 | .061 | .060 | .059 |
| | | 500 | .043 | .043 | .042 | .042 | .041 |
| | | 1000 | .032 | .032 | .031 | .031 | .030 |
| | 32 | 250 | .061 | .061 | .059 | .058 | .056 |
| | | 500 | .044 | .043 | .042 | .041 | .039 |
| | | 1000 | .031 | .031 | .030 | .029 | .028 |

Tabla A3

RMSE promedio de medias para no-invarianza en pesos

| Método | Número de grupos | Tamaño muestral | $\Delta\lambda = 0$ | $\Delta\lambda = .1$ | $\Delta\lambda = .2$ | $\Delta\lambda = .3$ | $\Delta\lambda = .4$ |
|--------------|------------------|-----------------|---------------------|----------------------|----------------------|----------------------|----------------------|
| AFC-MG | 4 | 250 | .094 | .094 | .094 | .094 | .094 |
| | | 500 | .067 | .067 | .067 | .067 | .067 |
| | | 1000 | .047 | .047 | .047 | .047 | .047 |
| | 8 | 250 | .095 | .095 | .095 | .095 | .095 |
| | | 500 | .067 | .067 | .067 | .067 | .067 |
| | | 1000 | .046 | .046 | .046 | .046 | .046 |
| | 16 | 250 | .095 | .095 | .095 | .095 | .095 |
| | | 500 | .067 | .067 | .067 | .067 | .067 |
| | | 1000 | .047 | .046 | .046 | .046 | .046 |
| | 32 | 250 | .094 | .094 | .094 | .094 | .094 |
| | | 500 | .067 | .067 | .067 | .067 | .067 |
| | | 1000 | .047 | .046 | .046 | .046 | .046 |
| Alineamiento | 4 | 250 | .095 | .095 | .095 | .095 | .095 |
| | | 500 | .065 | .065 | .065 | .065 | .065 |
| | | 1000 | .046 | .046 | .046 | .046 | .045 |
| | 8 | 250 | .095 | .095 | .095 | .095 | .095 |
| | | 500 | .066 | .066 | .067 | .066 | .066 |
| | | 1000 | .047 | .047 | .047 | .047 | .046 |
| | 16 | 250 | .095 | .095 | .095 | .095 | .094 |
| | | 500 | .065 | .065 | .065 | .065 | .064 |
| | | 1000 | .047 | .047 | .047 | .047 | .046 |
| | 32 | 250 | .098 | .099 | .100 | .102 | .188 |
| | | 500 | .067 | .067 | .067 | .067 | .066 |
| | | 1000 | .047 | .047 | .047 | .046 | .045 |

Tabla A4
 RMSE promedio de pesos para no-invarianza en intersecciones

| Método | Número de grupos | Tamaño muestral | $\Delta v = 0$ | $\Delta v = .1$ | $\Delta v = .2$ | $\Delta v = .3$ | $\Delta v = .4$ | $\Delta v = .5$ |
|--------------|------------------|-----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| AFC-MG | 4 | 250 | .043 | .043 | .043 | .043 | .043 | .043 |
| | | 500 | .031 | .031 | .031 | .031 | .031 | .031 |
| | | 1000 | .022 | .022 | .022 | .022 | .022 | .022 |
| | 8 | 250 | .041 | .041 | .041 | .041 | .041 | .041 |
| | | 500 | .029 | .029 | .029 | .029 | .029 | .029 |
| | | 1000 | .020 | .020 | .020 | .020 | .020 | .020 |
| | 16 | 250 | .041 | .041 | .041 | .041 | .041 | .041 |
| | | 500 | .029 | .029 | .029 | .029 | .029 | .029 |
| | | 1000 | .020 | .020 | .020 | .020 | .020 | .020 |
| | 32 | 250 | .040 | .039 | .039 | .039 | .039 | .039 |
| | | 500 | .028 | .028 | .028 | .028 | .028 | .028 |
| | | 1000 | .020 | .020 | .020 | .020 | .020 | .020 |
| Alineamiento | 4 | 250 | .054 | .054 | .054 | .054 | .054 | .054 |
| | | 500 | .039 | .039 | .039 | .039 | .039 | .039 |
| | | 1000 | .027 | .027 | .027 | .027 | .027 | .027 |
| | 8 | 250 | .056 | .056 | .056 | .056 | .056 | .056 |
| | | 500 | .038 | .038 | .038 | .038 | .038 | .038 |
| | | 1000 | .027 | .027 | .027 | .027 | .027 | .027 |
| | 16 | 250 | .060 | .060 | .060 | .060 | .060 | .060 |
| | | 500 | .040 | .040 | .040 | .040 | .040 | .040 |
| | | 1000 | .028 | .028 | .028 | .028 | .028 | .028 |
| | 32 | 250 | .075 | .075 | .075 | .075 | .075 | .075 |
| | | 500 | .046 | .046 | .046 | .046 | .046 | .046 |
| | | 1000 | .030 | .030 | .030 | .030 | .030 | .030 |

Tabla A5

RMSE promedio de varianzas para no-invarianza en intersecciones

| Método | Número de grupos | Tamaño muestral | $\Delta v = 0$ | $\Delta v = .1$ | $\Delta v = .2$ | $\Delta v = .3$ | $\Delta v = .4$ | $\Delta v = .5$ |
|--------------|------------------|-----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| AFC-MG | 4 | 250 | .139 | .139 | .139 | .139 | .139 | .139 |
| | | 500 | .099 | .099 | .099 | .099 | .099 | .099 |
| | | 1000 | .069 | .069 | .069 | .069 | .069 | .069 |
| | 8 | 250 | .141 | .141 | .141 | .141 | .141 | .141 |
| | | 500 | .099 | .099 | .099 | .099 | .099 | .099 |
| | | 1000 | .069 | .069 | .069 | .069 | .069 | .069 |
| | 16 | 250 | .144 | .144 | .144 | .144 | .144 | .144 |
| | | 500 | .100 | .100 | .100 | .100 | .100 | .100 |
| | | 1000 | .069 | .069 | .069 | .069 | .069 | .069 |
| | 32 | 250 | .142 | .142 | .142 | .142 | .142 | .142 |
| | | 500 | .099 | .099 | .099 | .099 | .099 | .099 |
| | | 1000 | .070 | .070 | .070 | .070 | .070 | .070 |
| Alineamiento | 4 | 250 | .072 | .072 | .072 | .072 | .072 | .072 |
| | | 500 | .050 | .050 | .050 | .050 | .050 | .050 |
| | | 1000 | .034 | .034 | .034 | .034 | .034 | .034 |
| | 8 | 250 | .074 | .074 | .074 | .074 | .074 | .074 |
| | | 500 | .050 | .050 | .050 | .050 | .050 | .050 |
| | | 1000 | .035 | .035 | .035 | .035 | .035 | .035 |
| | 16 | 250 | .082 | .082 | .082 | .082 | .082 | .082 |
| | | 500 | .052 | .052 | .052 | .052 | .052 | .052 |
| | | 1000 | .036 | .036 | .036 | .036 | .036 | .036 |
| | 32 | 250 | .106 | .106 | .106 | .106 | .106 | .106 |
| | | 500 | .062 | .062 | .062 | .062 | .062 | .062 |
| | | 1000 | .039 | .039 | .039 | .039 | .039 | .039 |

Apéndice C

Tablas resumen de potencia de ambos procedimientos

Tabla A6
Potencia para no-invarianza en pesos

| Método | Número de grupos | Tamaño muestral | $\Delta\lambda = .1$ | $\Delta\lambda = .2$ | $\Delta\lambda = .3$ | $\Delta\lambda = .4$ |
|--------------|------------------|-----------------|----------------------|----------------------|----------------------|----------------------|
| AFC-MG | 4 | 250 | .16 | .52 | .86 | .99 |
| | | 500 | .28 | .84 | .99 | 1 |
| | | 1000 | .58 | 1 | 1 | 1 |
| | 8 | 250 | .19 | .71 | .97 | 1 |
| | | 500 | .40 | .98 | 1 | 1 |
| | | 1000 | .78 | 1 | 1 | 1 |
| | 16 | 250 | .25 | .89 | 1 | 1 |
| | | 500 | .56 | 1 | 1 | 1 |
| | | 1000 | .96 | 1 | 1 | 1 |
| | 32 | 250 | .40 | .99 | 1 | 1 |
| | | 500 | .81 | 1 | 1 | 1 |
| | | 1000 | 1 | 1 | 1 | 1 |
| Alineamiento | 4 | 250 | .42 | .92 | 1 | 1 |
| | | 500 | .41 | .98 | 1 | 1 |
| | | 1000 | .35 | 1 | 1 | 1 |
| | 8 | 250 | .35 | .98 | 1 | 1 |
| | | 500 | .35 | 1 | 1 | 1 |
| | | 1000 | .28 | 1 | 1 | 1 |
| | 16 | 250 | .30 | 1 | 1 | 1 |
| | | 500 | .29 | 1 | 1 | 1 |
| | | 1000 | .26 | 1 | 1 | 1 |
| | 32 | 250 | .19 | 1 | 1 | .93 |
| | | 500 | .15 | 1 | 1 | 1 |
| | | 1000 | .11 | 1 | 1 | 1 |

Tabla A7
Potencia para no-invarianza en intersecciones

| Método | Número de grupos | Tamaño muestral | $\Delta v = .1$ | $\Delta v = .2$ | $\Delta v = .3$ | $\Delta v = .4$ | $\Delta v = .5$ |
|--------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| AFC-MG | 4 | 250 | .20 | .76 | .99 | 1 | 1 |
| | | 500 | .38 | .99 | 1 | 1 | 1 |
| | | 1000 | .77 | 1 | 1 | 1 | 1 |
| | 8 | 250 | .28 | .93 | 1 | 1 | 1 |
| | | 500 | .54 | 1 | 1 | 1 | 1 |
| | | 1000 | .94 | 1 | 1 | 1 | 1 |
| | 16 | 250 | .38 | 1 | 1 | 1 | 1 |
| | | 500 | .77 | 1 | 1 | 1 | 1 |
| | | 1000 | .99 | 1 | 1 | 1 | 1 |
| | 32 | 250 | .58 | 1 | 1 | 1 | 1 |
| | | 500 | .96 | 1 | 1 | 1 | 1 |
| | | 1000 | 1 | 1 | 1 | 1 | 1 |
| Alineamiento | 4 | 250 | 0 | .36 | .95 | 1 | 1 |
| | | 500 | 0 | .31 | .99 | 1 | 1 |
| | | 1000 | 0 | .29 | 1 | 1 | 1 |
| | 8 | 250 | 0 | .29 | .99 | 1 | 1 |
| | | 500 | 0 | .26 | 1 | 1 | 1 |
| | | 1000 | 0 | .21 | 1 | 1 | 1 |
| | 16 | 250 | 0 | .22 | 1 | 1 | 1 |
| | | 500 | 0 | .19 | 1 | 1 | 1 |
| | | 1000 | 0 | .12 | 1 | 1 | 1 |
| | 32 | 250 | 0 | .15 | 1 | 1 | 1 |
| | | 500 | 0 | .11 | 1 | 1 | 1 |
| | | 1000 | 0 | .06 | 1 | 1 | 1 |