

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

**COMPARACIÓN ENTRE MODELOS DE DATOS MULTINIVEL A
TRAVÉS DE ECUACIONES ESTRUCTURALES CON
ESTIMACIÓN BAYESIANA Y PEQUEÑAS VARIANZAS A
PRIORI EN LAS CARGAS FACTORIALES CRUZADAS**

Tesis sometida a la consideración de la Comisión del Programa de Estudios de Posgrado en
Estadística para optar al grado y título de Maestría Académica en Estadística

ANDRÉS FELIPE FERNÁNDEZ ARAUZ

Ciudad Universitaria Rodrigo Facio, Costa Rica

2020

Agradecimientos

El camino para la construcción y finalización de este documento estuvo acompañado de varias personas que, con sus aportes, ayudaron a moldear el trabajo presentado en esta tesis:

A Eiliana Montero Rojas, que como tutora de esta tesis me orientó en todas las áreas de investigación requeridas para desarrollar este documento, incluyendo los avances realizados en los cursos de modelos mixtos y modelos de ecuaciones estructurales.

A Ricardo Alvarado Barrantes, que como lector de tesis y primer profesor con el que tuve contacto en los cursos nivelatorios, me orientó en todos los aspectos relacionados a la implementación de la estimación Bayesiana.

A Guaner Rojas Rojas, que en su función de lector brindó valiosos aportes para mejorar los aspectos relacionados a las simulaciones presentadas en este documento.

A Johnny Madrigal Pana, que me introdujo en el área del análisis de factores latentes, particularmente por su colaboración para comprender el análisis factorial confirmatorio.

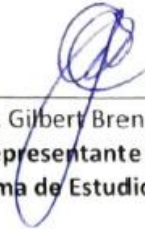
A Eiliana Montero Rojas por su motivación, y al Sistema de Estudios de Posgrado por el apoyo brindado, para asistir en la conferencia Modern Modeling Methods que se llevó a cabo en la Universidad de Connecticut, Estados Unidos, en mayo de 2015.

A James Clifton, de la Universidad de California, Merced, por sus valiosas contribuciones para el diseño de las simulaciones Montecarlo mediante el uso de MPlus, y a Sarah Depaoli, profesora asociada de Psicología Cuantitativa en la Universidad de California, Merced, por compartir algunos de los resultados de su trabajo en el área de modelación de factores latentes.

A Bengt Muthen, creador del software MPlus y autor de la propuesta de modelo BSEM sobre el cual se basa el desarrollo de esta tesis, y a Michael Hallquist, desarrollador de la librería MplusAutomation. A ambos por su disposición para discutir aspectos técnicos del uso de MPlus.

Finalmente, y no menos importante, a mi esposa Lucrecia, y mis dos hijos Felipe y Sara, por ser la inspiración para continuar avanzando en mi crecimiento personal y profesional.

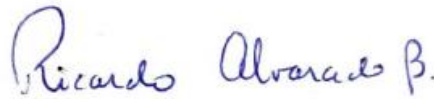
“Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar por el grado y título de Maestría Académica en Estadística.”



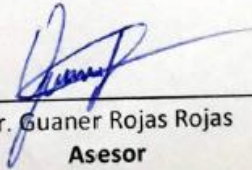
Dr. Gilbert Brenes Camacho
Representante del Decano
Sistema de Estudios de Posgrado



Dra. Eiliana Montero Rojas
Directora de Tesis



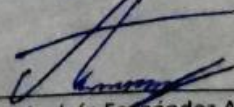
Dr. Ricardo Alvarado Barrantes
Asesor



Dr. Guaner Rojas Rojas
Asesor



M.Sc. Johnny Madrigal Pana
Director
Programa de Posgrado en Estadística



Andrés Fernández Aráuz
Candidato

Tabla de Contenidos

Portada.....	i
Agradecimientos	ii
Hoja de aprobación	iii
Resumen.....	vi
Abstract	vii
Lista de figuras	viii
Lista de tablas.....	ix
Lista de gráficos.....	xi
Licencia de publicación	xvi
1. Introducción	1
1.1. Problema	1
1.2. Objetivo General	2
1.3. Objetivos específicos.....	3
2. Estado de la cuestión sobre los modelos de ecuaciones estructurales multinivel	4
2.1. Análisis de factores.....	4
2.2. Ecuaciones estructurales.....	7
2.3. Ecuaciones estructurales multinivel.....	11
2.3.1. Análisis de factores multinivel.....	11
2.4. Métodos de estimación.....	16
2.4.1. Estimación frecuentista.....	16
2.4.2. Estimación Bayesiana	18
2.4.3. Enfoque Bayesiano en la estimación de modelos de ecuaciones estructurales.....	21
2.4.4. Integración de los modelos de ecuaciones estructurales multinivel con estimación Bayesiana	25
2.5. La discusión en el uso de las distribuciones a priori con varianzas pequeñas en el contexto de modelos de ecuaciones estructurales con estimación Bayesiana	29
3. Estado de la cuestión sobre el caso aplicado	35
3.1. Breve descripción de los factores asociados al rendimiento educativo	35
3.1.1. Factores del Estudiante	36
3.1.2. Factores Escolares	36
3.1.3. Factores institucionales y de política educativa.....	37
3.2. Estudios sobre factores asociados al rendimiento educativo	38
3.2.1. Diseños experimentales	39

3.2.2.	Otros estudios internacionales	41
3.2.3.	Revisión de literatura de Costa Rica.....	42
3.2.4.	Estudios que utilizan datos de PISA para Costa Rica.....	44
3.2.5.	Síntesis del capítulo.....	47
4.	Abordaje metodológico.....	49
4.1.	Evaluación simulada de la robustez del modelo ML-BSEM	49
4.1.1.	Descripción del modelo.....	49
4.1.2.	Especificación de las distribuciones a priori.....	51
4.1.3.	Diseño de las simulaciones.....	53
4.1.4.	Evaluación del ajuste del modelo.....	56
4.2.	Análisis de los datos de la evaluación PISA	58
4.2.1.	El Programa Internacional de Evaluación de Estudiantes (PISA)	58
4.2.2.	Marco muestral	63
4.2.3.	Diseño muestral	64
4.2.4.	Imputación de datos faltantes	66
4.2.5.	Estimación del modelo	67
4.2.5.1.	Constructos definidos para el nivel dentro de grupos	67
4.2.5.2.	Constructos definidos para el nivel entre grupos	71
5.	Resultados	73
5.1.	Resultados del estudio de simulación	73
5.1.1	Estabilidad de las estimaciones de los parámetros	73
5.1.1.1	Las distribuciones de las estimaciones de los parámetros	74
5.1.1.2	Raíz cuadrada del error cuadrático medio en los parámetros.....	77
5.1.2	Sesgos en las estimaciones de los parámetros	84
5.1.3	Bondad de ajuste de los modelos	91
5.2	Caso aplicado.....	94
6	Conclusiones.....	99
7	Referencias.....	102
8	Anexos	110
	Anexo 1 Distribuciones condicionales.....	110
	Anexo 2: Gráficos de dispersión del nivel dentro de grupos	112
	Anexo 3: Gráficos de dispersión del nivel entre grupos.....	127

Resumen

Debido a las bondades del enfoque Bayesiano para la estimación de modelos de ecuaciones estructurales, en la última década se ha desarrollado un nuevo enfoque con la intención de producir un análisis que refleje de una mejor manera las teorías del investigador y sus creencias a priori. Esto se hace al utilizar sistemáticamente distribuciones a priori informativas para los parámetros que no deberían ser estimados libremente de acuerdo con la teoría del investigador. En el análisis frecuentista tales parámetros son fijados como iguales a cero, pero fueran liberados y debieran ser estimados el modelo presentaría problemas de identificación y de estimación. Mediante el enfoque Bayesiano, por el contrario, se logra la identificación del modelo al impulsar el uso de varianzas a priori muy pequeñas para estos parámetros.

Si bien es reciente la discusión sobre el uso de varianzas pequeñas en las distribuciones a priori de parámetros que usualmente son fijados a ser cero, nada ha sido discutido acerca de las implicaciones de su aplicación en el contexto de ecuaciones estructurales para datos jerárquicos o multinivel. Por tales motivos, la presente investigación aborda esta discusión en el contexto de modelos de ecuaciones estructurales para datos multinivel con estimación Bayesiana al plantear modelos en los que fue liberada la estimación de parámetros en las cargas factoriales cruzadas. Para hacer esto, fueron establecidos tres niveles de información a introducir en la varianza de las cargas factoriales cruzadas: poco informativa, débilmente informativa e informativa.

El estudio de simulación permitió mostrar que, en el nivel inferior (o dentro de grupos) el uso de distribuciones a priori informativas en las cargas factoriales cruzadas mejora la precisión con la que son estimados los parámetros principales. En el nivel superior jerárquico (o entre grupos), a nivel general la estimación Bayesiana con varianzas a priori informativas en las cargas factoriales cruzadas presenta resultados muy superiores a los de las demás condiciones de estimación.

El ajuste global del modelo al comparar el uso de distribuciones a priori débilmente informativas y difusas tiende a ser similar; sin embargo, se refuerza el hecho de que cuando se tienen pocos datos, un pequeño número de grupos y valores bajos del ICC, el uso de distribuciones a priori muy informativas en las cargas factoriales cruzadas sí produce mejores ajustes del modelo global.

Se puede afirmar que, de forma consistente, el uso de distribuciones a priori con varianza pequeña o muy informativa en las cargas factoriales cruzadas produce mejores resultados bajo distintas condiciones de estimación de los modelos de ecuaciones estructurales multinivel, por lo que esta investigación aporta evidencia para que este enfoque pueda ser replicado en otras investigaciones que busquen obtener ganancias en la estimación de modelos de ecuaciones estructurales para datos multinivel con enfoque Bayesiano.

Abstract

In the last decade, the use of very small variance prior of cross-factor loadings has been discussed in the context of estimating confirmatory factor analysis models or structural equations with a Bayesian approach (Muthén, B., & Asparouhov, T, 2012, 2013)

Traditionally, researches that relies on confirmatory factor analysis establish cross-factor loadings as equal to zero, meaning that the particular indicator has no relation to the latent factor or construct for which the factor load has been established as exactly equal to zero.

However, the indicator variables are rarely perfectly pure construction indicators, especially in the study of social sciences, which implies that significant levels of association with multiple constructs can occur.

In fact, in the reflexive logic of factor analysis, latent factors are those that have an influence on the indicators, rather than the other way around. This means that establishing small cross-loads reflects the influence of the factor on the relevant part of the construction of the indicators, rather than the indicators having an impact on the nature of the factor itself.

For this reason, this research addresses this discussion in the context of Multilevel Bayesian Structural Equation Models (ML-BSEM) by proposing models in which the estimation of parameters in cross factor loadings will be released. Therefore, it is necessary to examine whether for this type of models the discussion described in the literature is maintained and if there are gains in the goodness of fit of the models by allowing that condition.

This will be done by evaluating the possible effects on the parameter estimates, the hypothesis tests associated with them, as well as the tests of goodness of fit of models, which may vary according to the definition of the prior distributions of the cross loadings, from little informative or diffuse to very informative, in the context of estimating Multilevel Bayesian Structural Equation Models. For this, in addition, robustness will be analyzed for different scenarios in which hierarchical models can operate, such as different group sizes at the high level, different sample sizes in the low level groups and different intensities of the intraclass correlation.

In general terms, the results show that the goodness of fit when comparing weakly informative priors and diffuse priors distributions for the cross-loadings tends to be similar, especially for large samples and ICC values; However, when there is little data, a small number of groups and low ICC values, the use of very informative prior distributions with small variance priors for the cross loadings does produce better fit.

Lista de figuras

Figura 1 Representación de un modelo confirmatorio de tres factores sin errores correlacionados	6
Figura 2 Modelo de ecuaciones estructurales con un factor latente endógeno	8
Figura 3 Representación de los dos enfoques para la estimación de análisis de factores multinivel	15
Figura 4 Modelo de generación de datos con coeficiente de correlación intra-clase de 0.05	49
Figura 5 Relación entre factores del modelo de ecuaciones estructurales a ser estimado con datos de la evaluación PISA 2012.....	71

Lista de tablas

Tabla 1 Descripción de las escalas de competencia en la evaluación matemática de PISA 2012	63
Tabla 2 Estadísticas descriptivas de la muestra de datos de PISA para Costa Rica según categoría. 2012	65
Tabla 3 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 1)	77
Tabla 4 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 2)	78
Tabla 5 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 3)	78
Tabla 6 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 4)	79
Tabla 7 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 5)	79
Tabla 8 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 6)	80
Tabla 9 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 1).....	81
Tabla 10 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 2).....	81
Tabla 11 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 3).....	82
Tabla 12 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 4).....	82
Tabla 13 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 5).....	83
Tabla 14 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 6).....	83
Tabla 15 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 1)	84

Tabla 16 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 2)	85
Tabla 17 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 3)	85
Tabla 18 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 4)	86
Tabla 19 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 5)	86
Tabla 20 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 6)	87
Tabla 21 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 1).....	87
Tabla 22 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 2).....	88
Tabla 23 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 3).....	88
Tabla 24 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 4).....	89
Tabla 25 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 5).....	89
Tabla 26 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro entre grupos, escenario 6).....	90
Tabla 27 Comparación de las estimaciones de las cargas factoriales de los factores latentes del modelo dentro de grupos.....	95
Tabla 28 Cargas factoriales cruzadas con estimaciones distintas del parámetro cero (Escenario 2).....	96
Tabla 29 Estimación de los parámetros estructurales del nivel dentro de grupos	97
Tabla 30 Comparación de las estimaciones de las cargas factoriales de los constructos del modelo entre grupos.....	98

Lista de gráficos

Gráfico 1 PPP-Value, escenario 1: Intervalo de credibilidad del P-value Predictivo a Posteriori, según información en la varianza de la distribución a priori de las Cargas Factoriales Cruzadas, Tamaño de la Muestra y Coeficiente de Correlación Intraclassa	92
Gráfico 2 PPP-Value, escenario 3: Intervalo de credibilidad del P-value Predictivo a Posteriori, según información en la varianza de la distribución a priori de las Cargas Factoriales Cruzadas, Tamaño de la Muestra y Coeficiente de Correlación Intraclassa	92
Gráfico 3 PPP-Value, escenario 6: Intervalo de credibilidad del P-value Predictivo a Posteriori, según información en la varianza de la distribución a priori de las Cargas Factoriales Cruzadas, Tamaño de la Muestra y Coeficiente de Correlación Intraclassa	93
Gráfico 4 Cargas factoriales directas, escenario 1: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	112
Gráfico 5 Cargas factoriales directas, escenario 2: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	112
Gráfico 6 Cargas factoriales directas, escenario 3: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	113
Gráfico 7 Cargas factoriales directas, escenario 4: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	113
Gráfico 8 Cargas factoriales directas, escenario 5: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	114
Gráfico 9 Cargas factoriales directas, escenario 6: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	114
Gráfico 10 Cargas factoriales cruzadas, escenario 1: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	115
Gráfico 11 Cargas factoriales cruzadas, escenario 2: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	115
Gráfico 12 Cargas factoriales cruzadas, escenario 3: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	116
Gráfico 13 Cargas factoriales cruzadas, escenario 4: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	116
Gráfico 14 Cargas factoriales cruzadas, escenario 5: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)	117

Gráfico 60 Varianza factorial, escenario 3: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)	140
Gráfico 61 Varianza factorial, escenario 4: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)	140
Gráfico 62 Varianza factorial, escenario 5: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)	141
Gráfico 63 Varianza factorial, escenario 6: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)	141



Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, Andrés Felipe Fernández Arauz, con cédula de identidad 1-1282-0400, en mi condición de autor del TFG titulado Comparación entre modelos de datos multinivel a través de Ecuaciones Estructurales con estimación Bayesiana y pequeñas varianzas a priori en las cargas factoriales cruzadas

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI NO *

*En caso de la negativa favor indicar el tiempo de restricción: _____ año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

INFORMACIÓN DEL ESTUDIANTE:

Nombre Completo: Andrés Felipe Fernández Arauz

Número de Carné: A41998 Número de cédula: 1-1282-0400

Correo Electrónico: affa17@gmail.com

Fecha: 30 de octubre de 2020 Número de teléfono: 8840-2711

Nombre del Director (a) de Tesis o Tutor (a): Eiliana Montero Rojas, PhD

ANDRES FELIPE FERNANDEZ ARAUZ (FIRMA)
Firmado digitalmente por ANDRES FELIPE FERNANDEZ ARAUZ (FIRMA)
Fecha: 2020.10.30 20:50:11 -06'00'

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

1. Introducción

1.1. Problema

Los modelos de ecuaciones estructurales con datos multinivel han ganado popularidad en los últimos años debido a la aparición de software estadísticos que permiten realizar estimaciones para datos que presentan estructuras jerárquicas. Si bien algunas de las principales limitaciones para su estimación han sido evaluadas y corregidas con el paso del tiempo y las mejoras en los algoritmos de estimación de diversos software, su reciente aparición aún plantea algunos retos pendientes.

La estimación Bayesiana vía Cadenas de Markov Monte Carlo da cuenta de una buena parte de la mejora en la estimación de los modelos de ecuaciones estructurales multinivel (donde la estimación Bayesiana produce estimaciones al menos tan buenas como las estimaciones frecuentistas, y en muchos casos superiores), pero su aplicación en el contexto de modelos de factores aún se ha visto restringida a muchos de los supuestos que usualmente eran establecidos en este tipo de modelos debido, precisamente, a los problemas de convergencia que se presentaban con los métodos de estimación frecuentista.

Particularmente, garantizar que el modelo esté sobre-identificado es una de las recomendaciones prácticas que por décadas se han aplicado en el análisis factorial (Brown, 2006). Un modelo está sobre-identificado si el número de parámetros desconocidos es menor al número de elementos conocidos. La diferencia entre ambos números de parámetros es por ende positiva y son los grados de libertad. Usualmente, para evitar problemas de convergencia del modelo, un mínimo de tres indicadores por cada variable latente es sugerido, además de establecer las cargas cruzadas como exactamente iguales a cero, todo con el objetivo de disminuir las posibilidades de no convergencia del modelo desde un enfoque frecuentista.

Pero en la última década se ha discutido el uso de varianzas muy pequeñas en las distribuciones a priori de las cargas factoriales cruzadas en el contexto de la estimación de modelos de análisis factorial confirmatorio o de ecuaciones estructurales con enfoque Bayesiano.

Tradicionalmente, las investigaciones que incursionan en el uso del análisis factorial confirmatorio establecen las cargas factoriales cruzadas como iguales a cero, queriendo indicar con esto que el indicador particular no tiene ninguna relación con el factor latente o constructo para el cual la carga factorial ha sido establecida como exactamente igual a cero.

Sin embargo, las variables indicadoras rara vez son indicadores de construcción perfectamente puros, especialmente en el estudio de las ciencias sociales, lo que implica que se pueden presentar niveles significativos de asociación con múltiples constructos.

De hecho, en la lógica reflexiva de los análisis factoriales, los factores latentes son los que tienen una influencia en los indicadores, en lugar de a la inversa. Esto quiere decir que establecer pequeñas

cargas cruzadas refleja la influencia del factor en la parte relevante de la construcción de los indicadores, en lugar de que los indicadores tengan un impacto en la naturaleza del factor en sí.

Por tales motivos, la presente investigación aborda esta discusión en el contexto de modelos de ecuaciones estructurales para datos multinivel con estimación Bayesiana al plantear modelos en los que fue liberada la estimación de parámetros en las cargas factoriales cruzadas. Por lo tanto, fue necesario examinar si para este tipo de modelos se mantiene la discusión descrita en la literatura y si existen ganancias en la bondad de ajuste de los modelos al permitir esa condición.

Esto fue realizado al evaluar los posibles efectos sobre las estimaciones de los parámetros, las pruebas de hipótesis asociadas a los mismos, así como las pruebas de bondad de ajuste de modelos, que pueden variar según la definición de las distribuciones a priori de las cargas factoriales cruzadas, desde poco informativas o difusas hasta muy informativas, en el contexto de estimación de modelos de ecuaciones estructurales para datos multinivel con estimación Bayesiana. Para esto, además, fue analizada la robustez para distintos escenarios en los cuales pueden operar los modelos jerárquicos, como distintos tamaños de grupos en el nivel alto, distintos tamaños de muestra en los grupos del nivel bajo y diferentes intensidades de la correlación intraclase.

La información resultante de este análisis fue la base para una segunda etapa de la investigación en la que se modelan variables provenientes de un test estandarizado de alfabetización matemática: la evaluación internacional PISA aplicada a estudiantes costarricenses en el año 2012. El estudio pretende generar una base para la toma de decisión sobre la manera de modelar los datos provenientes de la evaluación y la encuesta aplicada a estudiantes, que genera una base de datos con estructura jerárquica, tomando en cuenta sus características específicas.

Por lo tanto, este estudio pretende responder a la pregunta: ¿existen mejoras en la evaluación del ajuste de un modelo de ecuaciones estructurales para datos multinivel con estimación Bayesiana cuando se permite la definición de distribuciones a priori con varianzas pequeñas en las cargas factoriales cruzadas?

1.2. Objetivo General

Comparar el desempeño de modelos de ecuaciones estructurales para datos multinivel con estimación Bayesiana, según distintos escenarios de estimación de las cargas factoriales cruzadas, para distintos números de grupos, tamaños de grupos, y ante distintos grados de correlación intraclase.

1.3. Objetivos específicos

- Evaluar la estabilidad de los parámetros en la estimación de modelos de ecuaciones estructurales Bayesianos multinivel para distintas especificaciones de las distribuciones a priori de las cargas factoriales cruzadas.
- Evaluar los posibles sesgos en la estabilidad de los parámetros ante distintos escenarios de la estructura jerárquica de los datos.
- Determinar si el ajuste del modelo mejora al liberar la estimación de las cargas factoriales cruzadas.
- Determinar el modelo más adecuado para medir la asociación de distintas variables de los estudiantes y de sus centros educativos sobre el nivel de competencias en alfabetización matemática de los individuos, con base en los resultados obtenidos de los objetivos anteriores y de acuerdo a las características particulares de los datos de PISA.

2. Estado de la cuestión sobre los modelos de ecuaciones estructurales multinivel

Para poder comprender la escogencia del problema de investigación que se planteó en la pregunta de investigación enunciada en la introducción, resulta necesario brindar un marco detallado que permita contextualizar cómo se ha llegado hasta este momento en el debate de los modelos de ecuaciones estructurales Bayesianos multinivel. Por esto, se hará una introducción no exhaustiva a la teoría de análisis de factores (sección 2.1), los modelos de ecuaciones estructurales (sección 2.2), el análisis multinivel (sección 2.3) y la estimación Bayesiana (sección 2.4).

De esta forma, será hasta la sección 2.5 en la que se abordará por completo el estado actual de la discusión en torno al uso de distribuciones a priori con varianzas pequeñas en la estimación de parámetros que usualmente han sido pre fijados a un valor particular (usualmente media cero y varianza cero), como las cargas factoriales cruzadas, la covariación entre variables indicadoras, entre otras.

2.1. Análisis de factores

Desde su aparición hace más de un siglo, el análisis factorial se ha convertido en uno de los procedimientos estadísticos multivariantes más utilizados en la investigación aplicada, en una gran variedad de campos de aplicación (por ejemplo, la psicología, la educación, la sociología, la salud pública, entre otros). La intención fundamental del análisis factorial es determinar el número y la naturaleza de las variables latentes o factores que explican la variación y covariación entre un conjunto de medidas observadas, comúnmente conocidas como indicadores.

En concreto, un factor (también llamado variable latente o constructo) es una variable no observable que influye en más de una medida observada y que da cuenta de las correlaciones entre estas medidas observadas. En otras palabras, las medidas o variables observadas están interrelacionadas porque comparten un factor causal común (es decir, se encuentran influenciadas por el mismo constructo subyacente). Por lo tanto, el análisis factorial intenta una comprensión más parsimoniosa de la covariación entre un conjunto de indicadores debido a que el número de factores o constructos es menor que el número de variables observadas (Brown, 2006).

Las concepciones iniciales sobre análisis factorial provienen del modelo de factor común (Thurstone, 1947), que postula que cada indicador en un conjunto de medidas observadas es una función lineal de uno o más factores comunes y un factor único. Por lo tanto, el análisis factorial segmenta o parte la varianza de cada indicador en dos partes:

(1) la varianza común o varianza explicada por el factor latente, que es estimada sobre la base de la varianza compartida con otros indicadores en el análisis; y

(2) la varianza única, que es una combinación de la varianza que es específica para el indicador y la varianza del error aleatorio (es decir, error de medición en el indicador).

Hay dos tipos principales de análisis de factores: el análisis factorial exploratorio (EFA, por sus siglas en inglés) y el análisis factorial confirmatorio (CFA, por sus siglas en inglés; véase Joreskog 1969). Tanto el EFA como el CFA pretenden reproducir las relaciones observadas entre un grupo de indicadores con un conjunto más pequeño de variables latentes, pero se diferencian

fundamentalmente por el número y la naturaleza de las especificaciones a priori y restricciones hechas en el modelo de factores.

El análisis factorial exploratorio es un enfoque impulsado por los datos, de tal manera que no se hacen especificaciones en relación con el número de factores latentes o al patrón de las relaciones entre los factores comunes y los indicadores. Más bien, el investigador emplea el EFA como técnica exploratoria o descriptiva para determinar el número adecuado de factores comunes y de descubrir cuáles variables de medición son indicadores razonables de las diversas dimensiones latentes.

En el análisis factorial confirmatorio, el investigador especifica el número de factores y el patrón de relación entre el indicador y las cargas factoriales de antemano, así como otros parámetros. El factor de solución pre-especificado se evalúa en términos de lo bien que este reproduce la matriz de varianzas y covarianzas de la muestra en las variables medidas¹. Por lo tanto, a diferencia del EFA, el CFA requiere una sólida base empírica o conceptual para guiar la especificación y evaluación del modelo de factores.

La forma básica de un modelo CFA que expresa la relación entre las variables observadas (x), los factores latentes (ξ) y las varianzas únicas (ε) es la siguiente:

$$x = \Lambda\xi + \varepsilon$$

Donde Λ ($p \times q$) es una matriz de cargas factoriales (cuyos elementos se expresan con λ), ξ ($q \times 1$) es un vector aleatorio de factores latentes comunes y ε ($p \times 1$) es un vector aleatorio de errores de medición. Se asume que $\xi \sim N(0, \Phi)$ con una matriz simétrica definida positiva Φ ($q \times q$) que contiene las covariaciones entre los factores latentes; $\varepsilon \sim N(0, \Psi_\varepsilon)$ donde Ψ_ε ($p \times p$) es una matriz diagonal de varianzas únicas, y ξ no se encuentra correlacionada con ε .

El vector manifiesto x se distribuye como $N(0, \Sigma)$, donde la covarianza matricial de x es:

$$\Sigma = \Lambda\Phi\Lambda^T + \Psi_\varepsilon$$

Donde Σ ($p \times p$) es la matriz simétrica de varianzas y covarianzas de las p variables indicadoras o manifiestas. En este modelo, tradicionalmente los elementos de Λ , Φ y Ψ_ε pueden ser fijados de acuerdo con valores predeterminados.

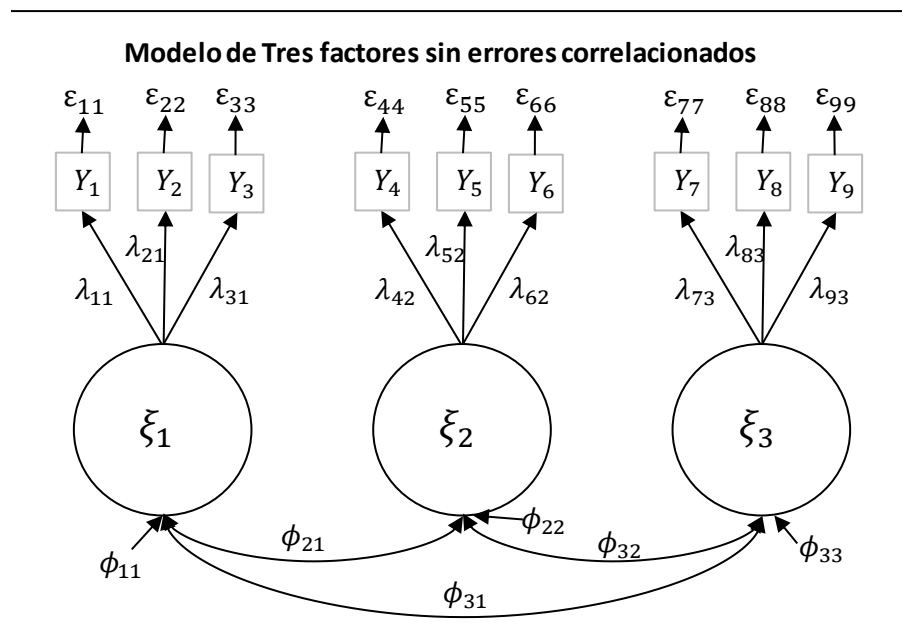
Por ejemplo, en un modelo de tres factores (ξ_1, ξ_2 y ξ_3) con tres variables indicadoras o manifiestas midiendo a cada factor (x_1, x_2, \dots, x_9), estas ecuaciones tendrían la siguiente forma:

¹ Este es el enfoque tradicional de análisis de factores, pero según fue explicado en el marco teórico, existen nuevos enfoques que no se basan en el ajuste de los datos a partir de la matriz de varianzas y covarianzas sino de las puntuaciones individuales de cada sujeto en la muestra.

$$\Lambda_{(9 \times 3)} = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \\ 0 & 0 & \lambda_{73} \\ 0 & 0 & \lambda_{83} \\ 0 & 0 & \lambda_{93} \end{pmatrix} \quad \Phi_{(3 \times 3)} = \begin{pmatrix} \phi_{11} & & \\ \phi_{21} & \phi_{22} & \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \quad \Psi_{\varepsilon(9 \times 9)} = \begin{pmatrix} \varepsilon_{11} & 0 & \dots & 0 \\ 0 & \varepsilon_{22} & & \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \varepsilon_{99} \end{pmatrix}$$

Gráficamente, estas matrices se representan de la siguiente manera:

Figura 1 Representación de un modelo confirmatorio de tres factores sin errores correlacionados



Frecuentemente, el análisis factorial confirmatorio es utilizado como precursor para los modelos de ecuaciones estructurales (SEM) que especifican relaciones estructurales (por ejemplo, regresiones) entre las variables latentes. Los SEM se pueden separar en dos grandes componentes:

- El modelo de medición, que especifica el número de factores, cómo se relacionan los indicadores con las variables latentes, y las relaciones entre los errores de los indicadores (es decir, el análisis factorial confirmatorio) y,
- El modelo estructural, que especifica cómo se relacionan unas con otras las distintas variables latentes (efectos directos o indirectos, ninguna relación, relaciones espurias).

De esta manera, un modelo de ecuaciones estructurales basa la relación estructural de las variables latentes en los resultados obtenidos del modelo de medición del análisis factorial confirmatorio, por lo que se encuentran estrechamente relacionados, y es claro por qué una solución de un CFA es una antesala muy útil a un modelo de ecuaciones estructurales, que intenta reproducir las relaciones

entre variables latentes con un set de parámetros estructurales más parsimonioso (es decir, menos parámetros) que el mismo CFA.

2.2. Ecuaciones estructurales

Los modelos de ecuaciones estructurales o SEM, por sus siglas en inglés (Structural Equations Models) se han hecho populares en las últimas décadas, en virtud de que representan una alternativa para explorar posibles planteamientos teóricos causales con datos que provienen de estudios observacionales o correlacionales (Montero, 2012).

Las raíces de los SEM se pueden trazar a partir de otras tres técnicas multivariadas más antiguas: el análisis de factores confirmatorio que proviene de la psicología, el análisis de trayectoria o de senderos (path analysis) que proviene de la biología y la genética, y los sistemas de ecuaciones simultáneas que se desarrollaron en economía. Un modelo SEM completo integra estas tres técnicas (Kaplan, 2009; Mulaik, 2009).

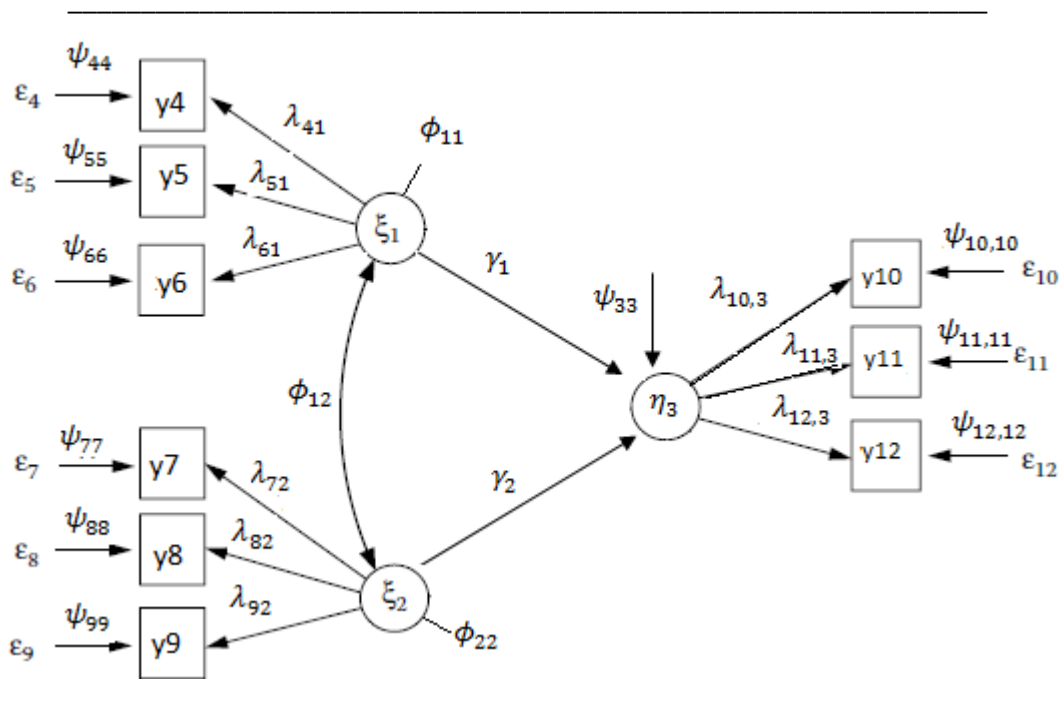
Los modelos de ecuaciones estructurales son una clase flexible de modelos que permiten la modelación compleja de datos multivariados correlacionados para evaluar las inter relaciones entre variables observadas y latentes. Esta clase de modelos subsume varios modelos estadísticos muy utilizados, como los modelos de regresión, análisis factorial, correlaciones canónicas y análisis de varianzas y covarianzas (Song & Lee, 2012).

Hoy día, debido a la fuerte demanda en varias disciplinas, hay más de una docena de paquetes de programas para estimar modelos SEM, como **AMOS** (un módulo del SPSS), **EQS6**, **LISREL**, **Stata** (gllamm), **R** (xxm), **Bugs** y **Mplus**. Cada software tiene distintas formas de especificar los SEM, pero la más ampliamente utilizada es la del modelo LISREL, que especifica un modelo de medición y un modelo estructural (Jöreskog & Sörbom, 1993).

A diferencia de los análisis estadísticos exploratorios, los modelos de ecuaciones estructurales se centran en probar hipótesis en modelos con ciertas restricciones sobre los parámetros previamente identificadas. El objetivo es probar si el modelo con estas restricciones ajusta bien a los datos con los cuáles se buscará realizar la prueba.

Una prueba clásica de modelos consiste en comparar la matriz de varianzas y covarianzas reproducida por las variables observadas contra la matriz de varianzas y covarianzas observada de los datos. Antes de presentar formalmente las ecuaciones involucradas en este proceso, resulta importante explicar los SEM desde una perspectiva gráfica. Esto se hará continuando con el modelo de tres factores de la sección anterior, sólo que en este caso los dos primeros factores latentes, llamados ξ_1 y ξ_2 , son factores determinantes del tercer factor, llamado η_3 , es decir, en este modelo hay dos factores latentes exógenos que tienen una relación directa sobre un factor latente endógeno, según se muestra en la siguiente figura:

Figura 2 Modelo de ecuaciones estructurales con un factor latente endógeno



Asociado a cada trayectoria causal directa hay un coeficiente estructural, que representa el efecto causal directo en la variable efecto. Este efecto representa en cuanto afecta una unidad de cambio de la variable causal sobre la variable efecto. En este caso, hay dos coeficientes estructurales que van desde los factores latentes exógenos hasta el factor latente endógeno, γ_1 y γ_2 .

Hay además nueve coeficientes de cargas factoriales, que modelan la relación entre cada variable indicadora (y_p) con su factor causal común². Las flechas para estas relaciones salen desde el factor latente exógeno y llegan hasta su variable indicadora, por lo que estas son también variables endógenas en el modelo. Estos coeficientes se representan con el símbolo λ , donde el primer número del subíndice que acompaña a λ se refiere a la variable y el segundo al factor en el que carga dicha variable.

Con el símbolo ϕ se representan los elementos de la matriz de varianzas y covarianzas de los factores latentes exógenos, mientras que el símbolo ψ representa la variable de error de cada variable endógena, sea latente o no.

Por último, con el símbolo ε se representa a los elementos de la diagonal de la matriz de varianzas y covarianzas de error entre las variables observadas, a lo cual es importante mencionar que todas las covarianzas son usualmente fijadas como cero.

² Al menos desde los enfoques clásicos, se trata de los 9 parámetros de cargas factoriales que son estimados libremente, ya que los restantes 18 parámetros de cargas factoriales cruzadas son fijados a ser cero.

El sistema de ecuaciones completo de este modelo está compuesto por diez ecuaciones (una para la variable latente endógena y nueve para las variables indicadoras endógenas), y es el siguiente:

$$\begin{aligned}
 \eta_3 &= \gamma_1 \xi_1 + \gamma_2 \xi_2 + \psi_{33} \varepsilon_3 \\
 y_4 &= \lambda_{41} \xi_1 + \psi_{44} \varepsilon_4 \\
 y_5 &= \lambda_{51} \xi_1 + \psi_{55} \varepsilon_5 \\
 y_6 &= \lambda_{61} \xi_1 + \psi_{66} \varepsilon_6 \\
 y_7 &= \lambda_{72} \xi_2 + \psi_{77} \varepsilon_7 \\
 y_8 &= \lambda_{82} \xi_2 + \psi_{88} \varepsilon_8 \\
 y_9 &= \lambda_{92} \xi_2 + \psi_{99} \varepsilon_9 \\
 y_{10} &= \lambda_{10,3} \eta_3 + \psi_{10,10} \varepsilon_{10} \\
 y_{11} &= \lambda_{11,3} \eta_3 + \psi_{11,11} \varepsilon_{11} \\
 y_{12} &= \lambda_{12,3} \eta_3 + \psi_{12,12} \varepsilon_{12}
 \end{aligned}$$

Este modelo puede ser reescrito de forma matricial de la siguiente manera:

$$\begin{bmatrix} \eta \\ y \end{bmatrix}_{10 \times 1} = A_{10 \times 10} \begin{bmatrix} \eta \\ y \end{bmatrix}_{10 \times 1} + \Gamma_{10 \times 2} \begin{bmatrix} \xi \\ x \end{bmatrix}_{2 \times 1} + \begin{bmatrix} \Delta_{1 \times 1} & 0 \\ 0 & \psi_{9 \times 9} \end{bmatrix}_{10 \times 2} \begin{bmatrix} \zeta \\ \epsilon \end{bmatrix}_{2 \times 1}$$

Si bien la nomenclatura presentada en la ecuación anterior aplica únicamente para el caso del modelo de tres factores, sirve de base para una introducción a los principales conceptos de los SEM. En los siguientes párrafos se explica con mayor detalle la forma matricial que adquieren estos modelos.

De acuerdo con Mulaik (2009) existen dos formas en las que puede ser expresado el modelo de ecuaciones que subyace a un modelo de ecuaciones estructurales con variables latentes. La primera de estas es la más usada para presentar ejemplos empíricos, por su facilidad de interpretación. Esta consiste en tratar a las perturbaciones de una manera separada que a las variables exógenas:

$$\begin{bmatrix} \eta \\ y \end{bmatrix} = A \begin{bmatrix} \eta \\ y \end{bmatrix} + \Gamma \begin{bmatrix} \xi \\ x \end{bmatrix} + \begin{bmatrix} \Delta & 0 \\ 0 & \psi \end{bmatrix} \begin{bmatrix} \zeta \\ \epsilon \end{bmatrix}$$

ó

$$\boldsymbol{\eta}^* = A \boldsymbol{\eta}^* + \boldsymbol{\Gamma}^* \boldsymbol{\xi}^* + \boldsymbol{\Delta}^* \boldsymbol{\epsilon}^*$$

Donde $\boldsymbol{\eta}^* = \begin{bmatrix} \eta \\ y \end{bmatrix}$ es el vector aleatorio de variables endógenas de tamaño $(m + p) \times 1$, con η siendo el vector de tamaño $(m \times 1)$ de variables latentes endógenas y y el vector de tamaño $(p \times 1)$ de variables manifiestas endógenas.

A es la matriz de coeficientes estructurales de tamaño $(m + p) \times (m + p)$ que relaciona las variables endógenas a otras variables endógenas

$\boldsymbol{\Gamma}^*$ es la matriz de coeficientes estructurales de tamaño $(m + p) \times (n + q)$ que relaciona las variables endógenas con las variables exógenas.

$\xi^* = \begin{bmatrix} \xi \\ x \end{bmatrix}$ es el vector aleatorio de variables endógenas de tamaño $(n + q) \times 1$, con ξ siendo el vector de variables latentes exógenas de tamaño $(n \times 1)$ y x el vector de tamaño $(q \times 1)$ de variables manifiestas exógenas.

ψ es una matriz diagonal de coeficientes estructurales, de tamaño $(p \times p)$, que relaciona las variables manifiestas endógenas a las perturbaciones exógenas.

ϵ es un vector de tamaño $(p \times 1)$ de variables aleatorias de perturbación en las variables manifiestas.

Δ es una matriz diagonal de coeficientes estructurales, de tamaño $(m \times m)$, que relaciona las m variables latentes endógenas (η) a las m perturbaciones (ζ).

ϵ es un vector de tamaño $(p \times 1)$ de variables aleatorias de perturbación en las variables manifiestas.

ϵ^* es el vector de tamaño $(m + p) \times 1$ de perturbaciones en las variables endógenas, tanto latentes como manifiestas.

La segunda forma de expresar en forma matricial los modelos de ecuaciones estructurales es tratando como variables exógenas tanto a las perturbaciones como a las variables manifiestas exógenas. Esto lleva a la siguiente representación matricial:

$$\begin{bmatrix} \eta \\ y \end{bmatrix} = A \begin{bmatrix} \eta \\ y \end{bmatrix} + \begin{bmatrix} \Gamma_\xi & \Gamma_x & \Gamma_\epsilon \end{bmatrix} \begin{bmatrix} \xi \\ x \\ \epsilon \end{bmatrix}$$

ó

$$\eta^* = A\eta^* + \Gamma^*\xi^*$$

Donde ξ^* es un vector de variables exógenas y de perturbaciones de tamaño $(n + q) \times 1$. No se hace ninguna distinción entre las perturbaciones sobre las variables latentes o sobre las variables manifiestas.

Esta forma es utilizada para introducir los diferentes métodos de estimación de los SEM, debido a que permite tratar a los γ 's y a los δ 's de una manera similar cuando se busca obtener las derivadas de las funciones de ajuste con respecto a sus elementos (James, Mulaik y Brett, 1982; Mulaik, 2009).

Es importante aclarar que aunque parezca que η^* aparece en ambos lados de la ecuación anterior, en realidad ninguna variable ocurre simultáneamente a ambos lados de la ecuación (aunque así parezca en la notación matricial), ya que los elementos de la diagonal de la matriz A son ceros (porque ninguna variable endógena puede ser causal de sí misma), lo que significa que ninguna variable endógena puede ocurrir a ambos lados de la ecuación.

Siguiendo a Mulaik (2009), la ecuación anterior puede expresarse en su forma reducida al aplicar una serie de operaciones algebraicas, como sigue:

$$\begin{aligned}\eta^* - A\eta^* &= \Gamma^* \xi^* \\ (I - A)\eta^* &= \Gamma^* \xi^*\end{aligned}$$

Haciendo $B = (I - A)$, se tiene que:

$$\begin{aligned}B\eta^* &= \Gamma^* \xi^* \\ \eta^* &= B^{-1}\Gamma^* \xi^*\end{aligned}$$

En esta última ecuación del modelo de ecuaciones expresado en su forma reducida, las variables endógenas son funciones matriciales únicamente de las variables exógenas.

2.3. Ecuaciones estructurales multinivel

Los enfoques de inferencia basados en la regresión por lo general requieren que los datos se recopilen de una muestra aleatoria simple en la que las observaciones son independientes. Sin embargo, muchos problemas de investigación en las ciencias sociales involucran el análisis de estructuras de datos jerárquicos, donde los sujetos se encuentran agrupados en uno o más niveles jerárquicos, como estudiantes agrupados dentro de escuelas o pacientes agrupados en hospitales.

En el contexto de datos jerárquicos, el término *nivel* denota diferentes componentes anidados. En el caso de los datos del caso práctico que fueron utilizados en esta Tesis posterior a las simulaciones, los sujetos son los estudiantes en edad de 15 años, y estos se encuentran agrupados o anidados en sus respectivos centros educativos. Esta es una estructura jerárquica en dos niveles: un primer nivel donde se encuentran los estudiantes y un segundo nivel donde están los centros educativos que contienen a estos estudiantes.

Utilizar enfoques de estimación basados en regresión que ignoren la naturaleza jerárquica de los datos (sea vía la desagregación al nivel de individuo o mediante la agregación al nivel de cluster o grupo) puede conducir a una variedad de problemas. La mayoría de estos problemas provienen de una posible violación al supuesto de observaciones independientes. La modelación multinivel es una técnica estadística que fue desarrollada para tomar en cuenta las dependencias que surgen en las estructuras de datos jerárquicos al modelar la variación entre las unidades de niveles altos y bajos (Raudenbush & Bryk, 2002).

2.3.1. Análisis de factores multinivel

Los modelos multinivel (MLM, siglas en inglés) y los modelos de ecuaciones estructurales (SEM, siglas en inglés) han evolucionado desde raíces conceptuales y metodológicas distintas. Los MLM tratan con el análisis de datos agrupados (por ejemplo, estudiantes dentro de centros educativos) y tienen como objetivo el descomponer la varianza observada entre los componentes *dentro de los grupos* y *entre los grupos*. Los SEM por su parte tratan con la modelación de medias y de las covarianzas en datos multivariados.

Aunque ambos se han desarrollado separadamente y para diferentes propósitos, los SEM y los MLM tienen importantes áreas en común ya que ambos enfoques incluyen variables latentes y efectos aleatorios para explicar las correlaciones entre las variables respuesta.

Estas dos clases de modelos se enfocan en preguntas diferentes y tienen distintas fortalezas y debilidades, y aunque algunos autores ya han estudiado la integración de ambos para aprovechar sus ventajas, la modelación estructural de datos con estructura multinivel es un área relativamente nueva de investigación metodológica (Metha & Neale, 2005; Bauer, 2003; Bentler & Liang, 2003; Curran, 2003; du Toit & du Toit, 2003; Rovine & Molennar, 2000; Depaoli & Clifton, 2015).

Los MLM se utilizan cuando la estructura de los datos es jerárquica con unidades elementales en el *nivel 1* anidadas en el *nivel 2* que a su vez podrían estar anidadas en el *nivel 3*, y así sucesivamente para *N niveles*. Las variables latentes, o efectos aleatorios, se interpretan como la heterogeneidad inobservada en los diferentes niveles que induce a la dependencia entre todas las unidades del nivel más bajo que pertenecen a una unidad de un nivel más alto. Las pendientes aleatorias representan la heterogeneidad entre grupos en la respuesta global y los coeficientes aleatorios representan la heterogeneidad en la relación entre la variable respuesta y las variables predictoras (Rabe-Hesketh y otros, 2004).

Los SEM se utilizan cuando las variables de interés no pueden ser medidas perfectamente. En su lugar, existe un set de ítems o variables indicadoras que reflejan un constructo hipotético. Las variables latentes o factores se interpretan como constructos que subyacen a las variables manifiestas y que induce dependencia entre estas.

La combinación de MLM y SEM ofrece a los investigadores la capacidad de responder una variedad de preguntas de investigación sofisticadas. Específicamente, los Modelos de Ecuaciones Estructurales Multinivel (ML-SEM, siglas en inglés) son útiles para testear las relaciones causales y tomar en cuenta el error de la medición de constructos con datos jerárquicos. Los ML-SEM pueden generalizarse a modelos con más de dos niveles de agrupamiento así como a datos longitudinales en los que los puntos temporales son un primer nivel agrupados en el segundo nivel de individuos. De esta forma, el enfoque ML-SEM es un marco muy general que proporciona la flexibilidad para estimar una gran variedad de modelos.

Un área común de aplicación de los ML-SEM son los modelos de medición multinivel, o análisis factorial confirmatorio multinivel, que brinda la posibilidad de especificar una estructura factorial distinta en cada nivel del modelo (Little, 2013; Kaplan, Kim and Kim, 2009; Diya, Li, Heede, Sermeus y Lesaffre, 2013).

ML-SEM se aplica también al contexto de análisis de trayectorias o modelos de mediación, siendo este último una trayectoria casual entre tres o más variables (Muthén B. , 1989); recientes avances en ML-SEM permiten la especificación de modelos de análisis de trayectorias con mediadores de nivel superior, un tipo de modelo que no puede ser estimado utilizando los enfoques tradicionales de modelación multinivel (Preacher, Zyphur, & Zhang, 2010).

Adicional al análisis de trayectoria multinivel, los ML-SEM pueden ser utilizados para combinar modelos de medición y modelos de trayectoria con datos multinivel, como los modelos multinivel de múltiples indicadores y múltiples causas (Finch y French, 2011; Morselli, Spine y Devos, 2012), y el modelo de covariación latente multinivel (Lüdtke, y otros, 2008)

Los ML-SEM pueden ser especificados de distintas maneras, es decir, no hay una única forma de lidiar con ML-SEM, así como tampoco hay una única forma de estimarlos. Los dos marcos más comunes para especificar ML-SEM son los modelos lineales generalizados latentes y mixtos (Rabe-Hesketh, Skrondal, & Pickles, 2004) y el enfoque en dos etapas (Muthén B. , 1994). El procedimiento más utilizado es el enfoque en dos etapas. En este caso, son especificados dos modelos de ecuaciones estructurales separadamente, uno para las matrices de covarianzas dentro de los grupos y otro para las matrices de covarianzas entre los grupos (Lee & Shi, 2001).

De acuerdo con este enfoque, para un modelo de factores en dos niveles para los sujetos j anidados en grupos k , el modelo se formula en términos de las matrices de covarianzas tanto *dentro de los grupos* (Σ_W) como *entre los grupos* (Σ_B). Para comprender esto, primero es necesario recordar la estructura factorial en un nivel. El modelo factorial convencional puede ser especificado como:

$$v_j = \beta\Lambda\eta_j \quad \text{con} \quad \eta_j \sim N(0, \Psi)$$

Para variables observadas o latentes continuas, se tiene que:

$$y_j^* = \beta\Lambda\eta_j + \varepsilon_j \quad \text{con} \quad \varepsilon_j \sim N(0, \Psi_\varepsilon)$$

Donde v_j y y_j^* son vectores que contienen los elementos de las variables indicadoras, β es un vector de interceptos, Λ es una matriz de cargas factoriales y η_j es la matriz de factores comunes. La estructura de covarianza de las respuestas latentes es:

$$\Sigma = Cov(y_j^*) = \Lambda\Phi\Lambda' + \Psi_\varepsilon$$

Para la estructura multinivel, en el enfoque en dos etapas el modelo factorial es especificado de la siguiente manera:

$$y_{jk}^* \sim N(\mu_k, \Sigma_W)$$

$$\mu_k \sim N(\mu, \Sigma_B)$$

Donde μ es el intercepto global y μ_k son interceptos específicos de cada grupo. La estructura de covarianzas de las respuestas latentes es ahora especificada con dos matrices de varianzas y covarianzas:

$$\Sigma_W = \Lambda^{(1)}\Phi^{(1)}\Lambda^{T(1)} + \Psi_\varepsilon^{(1)}$$

y

$$\Sigma_B = \Lambda^{(2)}\Phi^{(2)}\Lambda^{T(2)} + \Psi_\varepsilon^{(2)}$$

donde el superíndice (1) se utiliza para denotar a las variables y parámetros del nivel más bajo y (2) para denotar a sus contrapartes en el nivel de grupos.

El modelo factorial de dos factores puede ser expresado de una forma explícita usando una formulación en dos etapas con un modelo-dentro y un modelo-entre:

$$y_{jk}^* = \mu_k + \Lambda^{(1)}\eta_{jk}^{(1)} + \varepsilon_{jk}^{(1)}$$

Y

$$\mu_k = \mu + \Lambda^{(2)}\eta_{jk}^{(2)} + \varepsilon_{jk}^{(2)}$$

La primer ecuación para la variable respuesta y_{jk}^* representa un modelo de factores comunes que incluye interceptos aleatorios (μ_k) que varían entre grupos (k). La segunda ecuación representa un modelo de factores comunes para los interceptos aleatorios (μ_k).

Para Rabe-Hesketh y otros (2004), este método presenta algunas imitaciones. Por esta razón, proponen un marco conceptual distinto para modelar ecuaciones estructurales con multinivel, desarrollando los modelos lineales generalizados latentes y mixtos (GLLAMM, siglas en inglés) que consisten de un modelo respuesta y un modelo estructural.

El modelo respuesta es el mismo que el explicado en esta sección, pero generalizado para L niveles de agrupación. Condicional a las variables latentes, este modelo respuesta es un modelo lineal generalizado especificado vía un predictor lineal, una función de enlace y una distribución de la familia exponencial. Para un modelo con L niveles de agrupación y M_l variables latentes en el nivel $l > 1$, el predictor lineal tiene la siguiente forma:

$$v = x\beta + \sum_{l=2}^L \sum_{m=1}^{M_l} \eta_m^{(l)} Z_m^{(l)} \lambda_m^{(l)}$$

En la ecuación se omiten los subíndices del nivel de sujetos por simplicidad de presentación. Usualmente, el primer elemento de λ_m es restringido al valor de uno (1). Los elementos de x son covariables asociadas con los efectos fijos o coeficientes de regresión β .

La m -ésima variable latente M en el nivel l , $\eta_m^{(l)}$, es multiplicada por una combinación lineal $Z_m^{(l)} \lambda_m^{(l)}$ de covariables ($Z_m^{(l)}$), donde $\lambda_m^{(l)}$ son los parámetros (usualmente cargas factoriales).

Definiendo a $\boldsymbol{\eta}^{(l)} = (\eta_1^{(l)}, \eta_2^{(l)}, \dots, \eta_{M_l}^{(l)})$ como el vector de variables latentes en el nivel l para una unidad dada del nivel l con un vector correspondiente de covariables $\mathbf{z}^{(l)} = (\mathbf{z}_1^{(l)}, \mathbf{z}_2^{(l)}, \dots, \mathbf{z}_{M_l}^{(l)})$; y definiendo al vector de variables latentes de todos los niveles para una unidad del nivel 2 como $\boldsymbol{\eta} = (\boldsymbol{\eta}^{(2)}, \dots, \boldsymbol{\eta}^{(L)})$ con un vector de covariables $\mathbf{z} = (\mathbf{z}^{(2)}, \dots, \mathbf{z}^{(L)})$, la esperanza condicional de la respuesta y dados \mathbf{x} , \mathbf{z} y $\boldsymbol{\eta}$ es “enlazada” al predictor lineal v mediante una función de enlace $g(\cdot)$:

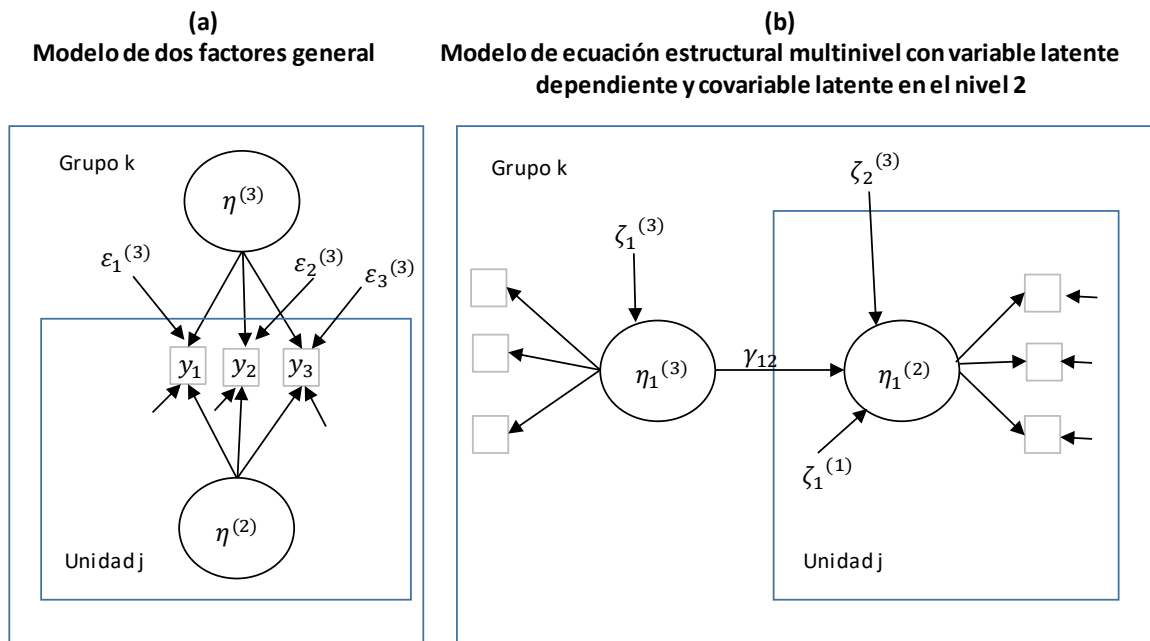
$$g(E[y|\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}]) = v$$

La especificación se completa al escoger una familia para la distribución condicional de la variable respuesta dados las variables latentes y las covariables. No hay variables latentes en el nivel 1, este nivel se reserva para el “error” de las distribuciones condicionales seleccionadas. En Rabe-Hesketh y otros (2004) se muestran algunas de estas distribuciones para diferentes tipos de variable respuesta³.

Este modelo permite la especificación de modelos de coeficientes aleatorios, modelos de medición, o ambos. El modelo estructural para L niveles tiene la forma tradicional para modelos de un nivel, pero es especificado para el vector η_j de todas las variables latentes para el sujeto j . Esto permite que variables latentes del primer nivel puedan ser regresadas en variables observadas o latentes de niveles superiores o del mismo nivel⁴. Este enfoque permite la especificación de modelos de coeficientes aleatorios con respuestas latentes en diferentes niveles.

El siguiente gráfico ilustra la diferencia entre el enfoque en dos etapas y el GLLAMM:

Figura 3 Representación de los dos enfoques para la estimación de análisis de factores multinivel



Una diferencia importante entre ambos enfoques es que en GLLAMM las ecuaciones para la estructura factorial terminan formando una única estructura que es estimada puntualmente para

³ Respuestas continuas, respuestas ordinales o dicotómicas, tiempos discretos, datos de conteo y duración en tiempo continuo, respuestas politómicas, y respuestas mixtas.

⁴ Sin embargo, GLLAMM no permite que las variables latentes sean regresadas en otras variables latentes u observadas que son aleatorias en el primer nivel, dado que esa especificación no tendría sentido interpretativo.

cada individuo, mientras que el enfoque general así como los SEM en un único nivel basan sus estimaciones en la estructura de la matriz de varianzas y covarianzas.

Si bien cada autor resalta las ventajas de su enfoque sobre el otro, en esencia, las diferencias entre el GLLAMM y el enfoque en dos etapas pueden ser consideradas en términos del formato de los datos. El marco GLLAMM requiere que los datos sean estructurados en un formato largo, de forma tal que todas las respuestas a los ítems estén contenidas en una única columna. Debido a que las respuestas de los individuos a cada ítem están apiladas, los ítems representan el nivel más bajo del modelo cuando se utiliza el marco GLLAMM. En contraste, el enfoque en dos etapas requiere que los datos se encuentren en formato amplio (multivariado) de forma tal que las respuestas a cada ítem estén contenidas en columnas separadas (el formato más usual). Consecuentemente, son las observaciones (y no los ítems) los que representan el nivel más bajo del modelo cuando se utiliza este enfoque. Debido a que el software que fue utilizado en esta investigación es *Mplus*, el enfoque de aplicación para la estimación de modelos multinivel es el de dos etapas.

2.4. Métodos de estimación

2.4.1. Estimación frecuentista

Para la estimación de los modelos de ecuaciones estructurales existen varios métodos como los mínimos cuadrados en dos etapas, mínimos cuadrados no ponderados, mínimos cuadrados generalizados, máxima verosimilitud, máxima verosimilitud robusta, mínimos cuadrados ponderados, entre otros (Joreskog y otros, 2000).

El primer método es no-iterativo y sirve para proporcionar las soluciones iniciales de los métodos iterativos, aunque sus soluciones pueden ser tomadas como la estimación final.

Los métodos iterativos proporcionan estimaciones basadas en la minimización de una función de ajuste, la cual modela las discrepancias entre las matrices de variancias y covarianzas observadas y estimadas. La función de ajuste del método de máxima verosimilitud es la más utilizada, y se expresa de la siguiente manera:

$$F_{ML} = \ln|S| - \ln|\Sigma| + \text{tr}(S)(\Sigma^{-1}) - p$$

Donde $|S|$ es el determinante de la matriz de varianzas y covarianzas observada, $|\Sigma|$ es el determinante de la matriz de varianzas y covarianzas estimada o predicha, p es el orden de la matriz "input" (el número de indicadores o variables manifiestas) y tr es una expresión para representar al operador de la traza una matriz.

El principio subyacente de la estimación vía máxima verosimilitud es encontrar las estimaciones de los parámetros del modelo que maximizan la probabilidad de observar los datos disponibles si los datos hubieran sido recolectados de la misma población otra vez. En otras palabras, la estimación por máxima verosimilitud tiene como propósito encontrar los valores de los parámetros que hacen que los datos observados sean más probables.

Los programas computacionales como **LISREL** (Jöreskog & Sörbom, 1993), **Mplus** (Muthén & Muthén, 1998-2015) y otros comienzan con un set inicial de parámetros conocidos como los valores iniciales (que pueden ser generados automáticamente por el software) y repetidamente refina esta estimación en un esfuerzo por reducir el valor de F_{ML} de la ecuación anterior (Brown, 2006). La convergencia del modelo se alcanza cuando el programa llega a un set de parámetros estimados que no pueden ser mejorados (es decir, que no pueden reducir F_{ML} más allá de lo ya alcanzado)⁵.

Una razón por la que la estimación vía máxima verosimilitud ha sido la más utilizada es porque posee propiedades estadísticas deseadas, como la habilidad de proveer errores estándar para cada uno de los parámetros estimados del modelo, que pueden ser utilizados para realizar pruebas de significancia estadística de los parámetros estimados y para determinar la precisión de estas estimaciones. Además, F_{ML} es utilizado en el cálculo de una gran cantidad de índices de bondad de ajuste (Brown, 2006).

Sin embargo, la estimación máximo verosímil tiene varios requerimientos que en algunos casos puede producir estimadores inadecuados, como varianzas negativas y coeficientes estandarizados menores a uno, y en el caso de modelos de ecuaciones estructurales multinivel puede enfrentar problemas de no convergencia, estimaciones sesgadas cuando el índice de correlación intraclase es bajo, cuando se dispone de un reducido número de observaciones o cuando se cuenta con un pequeño número de grupos en el nivel superior de la estructura multinivel (Depaoli y Clifton, 2015).

Además, algunos de los supuestos de la estimación máximo verosímil son:

- Tamaño de muestra grande (asintóticamente)
- Indicadores o variables manifiestas medidas en escalas continuas.
- La distribución de los indicadores es normal multivariada.

Aunque los parámetros efectivamente estimados mediante máxima verosimilitud pueden no verse afectados, la no normalidad podría resultar en errores estándar sesgados, con implicaciones en las pruebas de significancia y en la prueba bondad de ajuste chi cuadrado⁶. En este caso, se recomienda la utilización del método de máxima verosimilitud robusta (Brown, 2006). Si algunas de las variables manifiestas o indicatoras no son continuas, es recomendable utilizar mínimos

⁵ Los algoritmos de estimación más utilizados son el Algoritmo Newton-Raphson y los Métodos Cuasi-Newton (Mulaik, 2009).

⁶ Sin embargo, si la no normalidad es extrema la estimación máximo verosímil producirá estimaciones de los parámetros incorrectas debido a que el supuesto de linealidad en el modelo es inválido.

cuadrados ponderados, mínimos cuadrados ponderados robustos y mínimos cuadrados no ponderados.

La estimación multinivel mediante GLLAMM, el enfoque general y algunos otros enfoques basan sus estimaciones en máxima verosimilitud, aunque el GLLAMM puede adaptarse a métodos Monte Carlo (MCMC). De hecho, debido a la complejidad de los modelos SEM multinivel, frecuentemente se presentan problemas como estimaciones negativas para componentes de la matriz de varianzas y covarianzas estimadas (Hox & Maas, 2001; Lüdtke y otros., 2011) y estimaciones sesgadas de los parámetros (Li & Beretvas, 2013; Lüdtke y otros., 2008).

Dados los problemas que pueden presentarse con la estimación frecuentista, actualmente está ganando popularidad el uso de la estimación Bayesiana aplicados a una gran cantidad de modelos, dentro de los cuales se encuentran los modelos de ecuaciones estructurales.

Song & Lee (2012) definen claramente las ventajas de la estimación Bayesiana, basada en las observaciones aleatorias individuales brutas, sobre la estimación frecuentista, basada en la matriz de covarianza muestral:

- a) Primero, el desarrollo de métodos estadísticos está basado en las propiedades del primer momento de las observaciones individuales brutas que son más simples que las propiedades de segundo momento de la matriz de covarianza muestral.
- b) Segundo, se producen estimaciones directas de las variables latentes que no pueden ser obtenidas con los métodos clásicos.
- c) Tercero, directamente se modelan variables observadas con sus variables latentes a través de ecuaciones de regresión. De esta forma produce una interpretación más directa y se pueden utilizar las técnicas comunes de regresión para el análisis de datos
- d) Cuarto, además de la información que proporcionan los datos observados, el enfoque Bayesiano permite el uso de genuina información previa para producir mejores resultados.
- e) Quinto, el enfoque Bayesiano proporciona estadísticas más sencillas de evaluar para la bondad del ajuste y la comparación de modelos, así como otras estadísticas útiles como la media y los percentiles de la distribución posterior.
- f) Sexto, brinda resultados más confiables para muestras pequeñas.

Con el objetivo de evitar resultados no deseados o la no convergencia del modelo, en esta Tesis también fue utilizada la estimación Bayesiana, siguiendo el marco de referencia que se describe a continuación.

2.4.2. Estimación Bayesiana

El enfoque tradicional para analizar los modelos SEM es el del análisis de la estructura de covarianzas. Bajo este enfoque, la teoría estadística así como los algoritmos computacionales son desarrollados sobre la base de la matriz de covarianza muestral y su distribución asintótica. Este

enfoque funciona bien bajo algunos supuestos estándar, como por ejemplo que las observaciones aleatorias son normales, idéntica e independientemente distribuidas (Bentler & Liang, 2003). Como resultado, casi todos los desarrollos comerciales para modelar SEM fueron diseñados sobre la base de este enfoque. Sin embargo, en situaciones un poco más complejas que son comunes en la investigación aplicada -especialmente en las ciencias sociales- el enfoque de análisis de estructura de covarianza no es eficaz y puede tener problemas teóricos y computacionales (Song & Lee, 2012).

La estimación Bayesiana representa gran giro con respecto a los métodos frecuentistas de estimación de parámetros que han sido por varias décadas los más utilizados, particularmente por el uso de Cadenas de Markov Monte Carlo (MCMC, siglas en inglés) y el uso de una distribución a priori que aporta información al proceso de estimación y que no está presente en los métodos frecuentistas⁷.

Una de las diferencias fundamentales entre ambos métodos, además de los metodológicos, es la forma en que es vista la naturaleza de los parámetros poblacionales. La estimación clásica vía máxima verosimilitud asume que el vector de parámetros desconocidos no es aleatorio, lo que produce la estimación del parámetro poblacional en un único valor utilizando únicamente datos muestrales. Por el contrario, en el enfoque Bayesiano el vector de parámetros desconocidos se considera aleatorio con una distribución a priori y una distribución de densidad; es decir, en el paradigma Bayesiano el parámetro poblacional es estimado como una distribución de valores en lugar de un único número utilizando tanto datos muestrales como información de la distribución a priori que establece el investigador (Prees, 2003). De esta forma, los métodos Bayesianos combinan esta información a priori sobre la naturaleza del parámetro con la información que proveen los datos muestrales para estimar una distribución a posteriori.

Como la estimación Bayesiana no produce un único valor del parámetro sino una distribución aleatoria del mismo, en la práctica se calcula la media, la mediana o la moda de esta distribución a posteriori cuando es requerido el reporte de un valor. Adicionalmente, las desviaciones estándar y los intervalos de densidad para los parámetros del modelo pueden ser calculados a partir de esta distribución a posteriori.

Un elemento clave en la estadística Bayesiana es la especificación de la distribución a priori para cada parámetro del modelo. Estos valores a priori pueden ser de dos tipos: las distribuciones a priori informativas son usualmente tomadas de la investigación y el conocimiento previo y pueden especificar de una manera precisa el rango en el que podría ubicarse el parámetro buscado, mediante el establecimiento de pequeñas varianzas. Por el contrario, las distribuciones a priori no informativas o difusas no se basan en investigación previa, sino que son seleccionadas deliberadamente con el objetivo de no imponer grandes restricciones a la distribución a posteriori del parámetro, debido a que no se cuenta con información útil para limitarlo en cierta medida, y por lo tanto son caracterizadas por el uso de grandes varianzas.

Al utilizar varianzas de gran tamaño, el investigador reconoce la falta de información referente a su conocimiento sobre la distribución a posteriori, y permite que la estimación del

⁷ Es importante aclarar que aunque MCMC se utiliza frecuentemente en el contexto Bayesiano, no es un método inherentemente Bayesiano, sino que es un enfoque algorítmico para remuestreo que es utilizado también, aunque en menor medida, en la estimación máxima verosímil (ML) y máxima verosimilitud restringida (REML).

parámetro sea más influenciada por la información que proveen los datos muestrales (la verosimilitud) que por la información a priori (Finch y otros, 2014).

Una de las ventajas de los Métodos Bayesianos es que, a diferencia de la Máxima Verosimilitud, no impone ningún supuesto acerca de la distribución de los datos. Así, la determinación del intervalo de credibilidad se puede escoger sin preocuparse incluso de si los datos provienen de distribuciones sesgadas o muy alejadas de la típica distribución normal.

Otra ventaja de la estadística Bayesiana es que provee estimaciones de los parámetros más precisas en los casos de muestras pequeñas con respecto a lo obtenido por Máxima Verosimilitud, y que no presenta los problemas de convergencia en modelos complejos que sí ocurren con los métodos frecuentistas, además de que puede ser igualmente utilizado en casos en los que la estimación frecuentista también funciona.

Para ilustrar la estadística Bayesiana aplicada a los modelos de ecuaciones estructurales, se asume que \mathbf{M} es un modelo SEM con un vector de parámetros desconocidos $\boldsymbol{\theta}$, y que el set de datos observados se denota por $\mathbf{Y} = (y_1, \dots, y_n)$ con tamaño de muestra n .

En el enfoque Bayesiano, $\boldsymbol{\theta}$ es considerado aleatorio con una distribución (llamada distribución a priori) y una función de densidad asociada, $p(\boldsymbol{\theta}|\mathbf{M})$.

Sea $p(\mathbf{Y}, \boldsymbol{\theta}|\mathbf{M})$ la función de densidad probabilística de la distribución conjunta de \mathbf{Y} y $\boldsymbol{\theta}$ dado \mathbf{M} . El comportamiento de $\boldsymbol{\theta}$ dados los datos es descrito completamente por la distribución condicional de $\boldsymbol{\theta}$ dado \mathbf{Y} . A esta distribución condicional se le llama distribución a posteriori de $\boldsymbol{\theta}$.

Sea $p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{M})$ la función de densidad de la distribución a posteriori. De acuerdo con el Teorema de Bayes, e introduciendo el modelo SEM (\mathbf{M}), la siguiente identidad es cierta:

$$p(\mathbf{Y}, \boldsymbol{\theta}|\mathbf{M}) = p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{M})p(\boldsymbol{\theta}|\mathbf{M}) = p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{M})p(\mathbf{Y}|\mathbf{M})$$

Como $p(\mathbf{Y}|\mathbf{M})$ no depende de $\boldsymbol{\theta}$ se puede tratar como una constante para un \mathbf{Y} fijo, y entonces se tiene:

$$\log p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{M}) \propto \log p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{M}) + \log p(\boldsymbol{\theta}|\mathbf{M})$$

Donde $p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{M})$ es la función de verosimilitud y $p(\boldsymbol{\theta}|\mathbf{M})$ es la función de densidad a priori. La función de densidad a posteriori, $p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{M})$, depende de ambas funciones.

Como la función de verosimilitud depende de los datos y la función de densidad a priori no, entre más grande sea el tamaño de la muestra de los datos más se va a acercar la estimación Bayesiana a la estimación frecuentista, dado que la función a priori juega un rol menos importante. Por el contrario, entre más pequeño sea el tamaño de muestra la función a priori juega un rol más importante que la función de verosimilitud de los datos. Esta es de hecho una de las ventajas de la estimación Bayesiana sobre la frecuentista, y por lo tanto, la escogencia de la distribución a priori es un tema fundamental en la teoría Bayesiana.

2.4.3. Enfoque Bayesiano en la estimación de modelos de ecuaciones estructurales

En esta sección se explica la teoría básica de estimación de un modelo de ecuaciones estructurales general desde un enfoque Bayesiano, siguiendo lo expuesto por Lee (2007). Un modelo de ecuaciones estructurales se compone de un modelo de medición y una ecuación estructural. El modelo de medición se define de la siguiente manera:

$$y_i = \Lambda \omega_i + \epsilon_i$$

Donde y_i es un vector aleatorio de los valores observados, Λ es una matriz de cargas factoriales y ϵ_i es un vector aleatorio de errores de medición independiente de ω_i . Se asume que ϵ_i se distribuye $N[0, \Psi_\epsilon]$ donde Ψ_ϵ es una matriz diagonal.

Sea $\omega_i = (\eta_i^T, \xi_i^T)^T$ una partición de ω_i en $q_1 \times 1$ vectores latentes dependientes de η_i y en $q_2 \times 1$ vectores latentes independientes de ξ_i . La ecuación estructural para evaluar la relación entre η_i y ξ_i está dada por:

$$\eta_i = \mathbf{A}\eta_i + \mathbf{\Gamma}\xi_i + \delta_i$$

Donde \mathbf{A} y $\mathbf{\Gamma}$ son matrices de parámetros desconocidos de los coeficientes de regresión y δ_i es un vector aleatorio de errores de medición. Al igual que en el análisis factorial, se asume que ξ_i se distribuye como $N[0, \Phi]$ y que δ_i se distribuye como $N[0, \Psi_\delta]$, donde Ψ_δ es una matriz diagonal y, ξ_i y δ_i son independientes.

Ahora, haciendo $\mathbf{Y} = (y_1, \dots, y_n)$ y $\mathbf{\Omega} = (\omega_1, \dots, \omega_n)$, y definiendo a θ como el vector de parámetros desconocidos en $\Lambda, \Psi_\epsilon, \mathbf{A}, \mathbf{\Gamma}, \Phi$ y Ψ_δ , para la estimación Bayesiana, se replican los datos observados \mathbf{Y} con la matriz de variables latentes $\mathbf{\Omega}$ considerando la distribución posterior conjunta $[\theta, \mathbf{\Omega} | \mathbf{Y}]$. Un número suficientemente grande de observaciones de esta distribución a posteriori es obtenido mediante *muestreo de Gibbs*⁸, que en la iteración $(j+1)$ realiza lo siguiente:

- a) Generar $\mathbf{\Omega}^{(j+1)}$ de $p(\mathbf{\Omega} | \theta^{(j)}, \mathbf{Y})$,
- b) Generar $\theta^{(j+1)}$ de $p(\theta | \mathbf{\Omega}^{(j+1)}, \mathbf{Y})$

La derivación de la distribución condicional $p(\mathbf{\Omega} | \theta, \mathbf{Y})$ (punto "a" del proceso anterior) está basado en la definición del modelo y de las propiedades de distribución de los vectores aleatorios y_i y ω_i . Para $i = 1, \dots, n$, los ω_i son mutuamente independientes y los y_i también son mutuamente independientes dados (ω_i, θ) . Entonces, se tiene la siguiente ecuación para la distribución condicional $p(\mathbf{\Omega} | \theta, \mathbf{Y})$:

$$p(\mathbf{\Omega} | \theta, \mathbf{Y}) = \prod_{i=1}^n p(\omega_i | y_i, \theta) \propto \prod_{i=1}^n p(\omega_i | \theta) p(y_i | \omega_i, \theta)$$

⁸ Es un caso especial del algoritmo de Metropolis-Hastings con un componente donde se usa como densidad propuesta la distribución a posteriori condicional completa.

Lee Sik-Yum (2007) muestra que la distribución condicional de ω_i dados $(y_i, \boldsymbol{\theta})$ se distribuye como una Normal con los siguientes parámetros:

$$(\omega_i | y_i, \boldsymbol{\theta}) \sim N \left[(\boldsymbol{\Sigma}_\omega^{-1} + \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1} \boldsymbol{\Lambda})^{-1} \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1} y_i, (\boldsymbol{\Sigma}_\omega^{-1} + \boldsymbol{\Lambda}^T \boldsymbol{\Psi}_\epsilon^{-1})^{-1} \right]$$

Con

$$\boldsymbol{\Sigma}_\omega = \begin{bmatrix} \mathbf{A}_0^{-1} (\boldsymbol{\Gamma} \boldsymbol{\Phi} \boldsymbol{\Gamma}^T + \boldsymbol{\Psi}_\delta) \mathbf{A}_0^{-T} & \mathbf{A}_0^{-1} \boldsymbol{\Gamma} \boldsymbol{\Phi} \\ \boldsymbol{\Phi} \boldsymbol{\Gamma}^T \mathbf{A}_0^{-1} & \boldsymbol{\Phi} \end{bmatrix}$$

Por otra parte, la distribución condicional de $\boldsymbol{\theta}$ dados $(\boldsymbol{\Omega}, \mathbf{Y})$ para la iteración señalada en el punto b), es proporcional a $p(\boldsymbol{\theta})p(\mathbf{Y}, \boldsymbol{\Omega} | \boldsymbol{\theta})$, por lo que resulta necesario establecer la distribución a priori $p(\boldsymbol{\theta})$ de los parámetros desconocidos incluidos en $\boldsymbol{\theta}$ ($\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon, \mathbf{A}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}$ y $\boldsymbol{\Psi}_\delta$).

Definiendo $\boldsymbol{\theta}_y$ como los parámetros desconocidos en $\boldsymbol{\Lambda}$ y en $\boldsymbol{\Psi}_\epsilon$ que se asocian con el modelo de medición, y definiendo $\boldsymbol{\theta}_\omega$ como los parámetros desconocidos en $\mathbf{A}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}$ y $\boldsymbol{\Psi}_\delta$ asociados con el modelo estructural que relaciona las variables latentes, es normal asumir que las distribuciones a priori de cada vector de parámetros es independiente, es decir, $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_y)p(\boldsymbol{\theta}_\omega)$. Combinando esta información con otros resultados (Lee, 2007), se tiene que:

$$p(\boldsymbol{\theta}_y, \boldsymbol{\theta}_\omega | \mathbf{Y}, \boldsymbol{\Omega}) \propto [p(\mathbf{Y} | \boldsymbol{\Omega}, \boldsymbol{\theta}_y)p(\boldsymbol{\theta}_y)][p(\boldsymbol{\Omega} | \boldsymbol{\theta}_\omega)p(\boldsymbol{\theta}_\omega)]$$

Donde ambos términos multiplicativos pueden ser tratados separadamente como las distribuciones marginales condicionales de $\boldsymbol{\theta}_y$ y $\boldsymbol{\theta}_\omega$.

La distribución marginal condicionada de $\boldsymbol{\theta}_y$ es $p(\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\epsilon | \mathbf{Y}, \boldsymbol{\Omega})$ a la cual se le asigna las siguientes distribuciones a priori: sean $\psi_{\epsilon k}$ y Λ_k el elemento de la diagonal "k" de $\boldsymbol{\Psi}_\epsilon$ y la k-ésima fila de $\boldsymbol{\Lambda}$ respectivamente:

$$\begin{aligned} \psi_{\epsilon k}^{-1} &\sim \text{Gamma}(\alpha_{0\epsilon k}, \beta_{0\epsilon k}) \\ [\Lambda_k | \psi_{\epsilon k}] &\sim N[\Lambda_{0k}, \psi_{\epsilon k} H_{0yk}] \end{aligned}$$

Con H_{0yk} siendo una matriz definida positiva. Más adelante en la especificación del modelo que fue ajustado con los datos se explicará la escogencia de los hiperparámetros para todas las distribuciones a priori⁹.

Por otra parte, la distribución marginal condicionada de $\boldsymbol{\theta}_\omega$ es proporcional a $p(\boldsymbol{\Omega} | \boldsymbol{\theta}_\omega)p(\boldsymbol{\theta}_\omega)$, para el cuál se puede demostrar (Lee, 2007) que la distribución marginal condicional es:

$$p(\boldsymbol{\Omega} | \boldsymbol{\theta}_\omega)p(\boldsymbol{\theta}_\omega) = [p(\boldsymbol{\Omega}_1 | \boldsymbol{\Omega}_2, \mathbf{A}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_\delta)p(\mathbf{A}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_\delta)][p(\boldsymbol{\Omega}_2 | \boldsymbol{\Phi})p(\boldsymbol{\Phi})]$$

Donde las distribuciones condicionales de $(\mathbf{A}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_\delta)$ y de $\boldsymbol{\Phi}$ pueden ser tratadas separadamente.

⁹ Para claridad, las distribuciones gamma, gamma invertida, Wishart y Wishart invertida, así como sus características se presentan en el anexo 1.

La distribución a priori para Φ es

$$\Phi^{-1} \sim \text{Wishart}[R_0, \rho_0]$$

Con R_0 siendo una matriz definida positiva.

Por último, siguiendo una lógica similar para la escogencia de las distribuciones a priori para Λ_k y $\psi_{\epsilon k}$, y haciendo $\Lambda_\omega = (\mathbf{A}, \mathbf{\Gamma})$, se tiene lo siguiente:

$$\begin{aligned} \psi_{\delta k}^{-1} &\sim \text{Gamma}(\alpha_{0\delta k}, \beta_{0\delta k}) \\ [\Lambda_{\omega k} | \psi_{\delta k}] &\sim N[\Lambda_{0\omega k}, \psi_{\delta k} H_{0\delta k}] \end{aligned}$$

2.4.3.1. Tratamiento de variables categóricas en el modelo

El modelo recién expuesto muestra la forma básica en la que puede estimarse un modelo de ecuaciones estructurales mediante el enfoque Bayesiano, cuando todas las variables observadas o manifiestas que son utilizadas para la conformación de las variables latentes del modelo de medición pueden ser tratadas como continuas.

Sin embargo, debido al diseño de los cuestionarios y dada la naturaleza de las ciencias sociales, los datos a menudo provienen de variables ordenadas categóricas que incluyen observaciones de una forma discreta. El cuestionario que fue analizado en esta investigación no escapa a este diseño.

Un método comúnmente utilizado para analizar este tipo de datos es tratar los valores asignados dentro de la variable categórica como continuos, y provenientes de una distribución normal. Este enfoque podría no conducir a problemas serios si los histogramas de las observaciones son simétricos y con la frecuencia más alta en el centro, en cuyo caso se puede asumir normalidad multivariada.

Sin embargo, en algunos casos los sujetos que toman el cuestionario seleccionan categorías en las colas del ordenamiento de la variable, por lo que los histogramas correspondientes son asimétricos o bimodales, y tratar estas variables como normales puede llevar a conclusiones erróneas (ver Olsson 1979 y Lee 1990).

Según (Song & Lee, 2012) un mejor enfoque para analizar este tipo de datos discretos es tratarlos como observaciones que provienen de una distribución normal latente continua con una especificación de umbral. En el enfoque Bayesiano, la idea básica al utilizar las variables categóricas es tratar las mediciones continuas latentes subyacentes como datos faltantes aleatorios, y aumentarlos con los datos observados en el análisis a posteriori. Utilizando esta estrategia de aumento de datos, el modelo que se basa en el set de datos completo se convierte en uno con variables continuas. En la estimación, secuencias de las observaciones de los parámetros estructurales, de las variables latentes y de los umbrales son simuladas de la distribución posterior conjunta vía algoritmos Monte Carlo (MCMC). Siguiendo a Lee (2007) y Song & Lee (2012) la especificación opera de la siguiente manera:

La ecuación de medición para un vector aleatorio observado v_i , de tamaño $p \times 1$, es la siguiente:

$$v_i = \mu + \Lambda\omega_i + \epsilon_i, \quad i = 1, \dots, n$$

Sea $\eta_i(q_1 \times 1)$ y $\xi_i(q_2 \times 1)$ los subvectores latentes de ω_i con la siguiente ecuación estructural asociada:

$$\eta_i = \Pi\eta_i + \Gamma\xi_i + \delta_i$$

Haciendo $\Lambda_\omega = (\Pi, \Gamma)$, la ecuación anterior puede ser reescrita como:

$$\eta_i = \Lambda_\omega\omega_i + \delta_i$$

Sea $v = (x^T, y^T)^T$, donde $x = (x_1, \dots, x_r)^T$ es un subconjunto de variables cuyas mediciones continuas exactas son observables, mientras que $y = (y_1, \dots, y_s)^T$ es el restante subconjunto de variables tal que las mediciones continuas correspondientes son inobservables.

La información asociada con y es dada por un vector categórico observable $z = (z_1, \dots, z_s)^T$. Cualquier variable latente en η o ξ puede ser relacionada con variables observadas ya sea en x o en z . Esto significa que cualquier variable latente puede tener variables observadas continuas o categóricas en sus indicadores.

La relación entre y y z es definida por un set de umbrales, de la siguiente manera:

$$z = \begin{cases} z_1 \\ \vdots \\ z_s \end{cases} \text{ si } \begin{cases} \alpha_{1,z_1} \leq y_1 < \alpha_{1,z_1} + 1, \\ \vdots \\ \alpha_{s,z_s} \leq y_s < \alpha_{s,z_s} + 1 \end{cases}$$

Donde para $k = 1, \dots, s$, z_k es un valor entero en $\{0, 1, \dots, b_k\}$, y $\alpha_{k,0} < \alpha_{k,1} < \dots < \alpha_{k,b_k} < \alpha_{k,b_k+1}$

Para la k -ésima variable, hay $b_k + 1$ categorías definidas por los umbrales desconocidos $\alpha_{k,j}$. Los valores enteros $\{0, 1, \dots, b_k\}$ de z_k son utilizados para especificar las categorías que contienen los elementos correspondientes en y_k . Además, para evitar problemas de identificación Song & Lee (2012) sugieren fijar los umbrales en las colas, $\alpha_{k,1}$ y α_{k,b_k} , con valores pre asignados, lo que implica seleccionar medidas de locación y dispersión para y_k . La forma más común de hacerlo es utilizando las frecuencias observadas y la distribución normal estándar, $N[0,1]$: para todo k se puede fijar $\alpha_{k,1} = \Phi^{*-1}f_{k,1}^*$ y $\alpha_{k,b_k} = \Phi^{*-1}f_{k,b_k}^*$, donde $\Phi^*(\cdot)$ es la función de distribución $N[0,1]$, $f_{k,1}^*$ es la frecuencia de la primer categoría y f_{k,b_k}^* es la frecuencia acumulada de la categoría con $z_k < b_k$. Para SEM lineales esta restricción implica que la media y la varianza de la variable continua subyacente y_k son 0 y 1, respectivamente.

Tomando en consideración tanto las variables continuas como las categóricas para la estimación del modelo SEM Bayesiano, se utiliza la estrategia de aumento de datos descrita anteriormente. Sea

$X = (x_1, \dots, x_n)$ la matriz de datos de variables continuas y $Z = (z_1, \dots, z_n)$ la matriz de datos con variables categóricas. Sea $Y = (y_1, \dots, y_n)$ la matriz de mediciones continuas y $\Omega = (\omega_1, \dots, \omega_n)$ la matriz de variables latentes. Los datos observados $[X, Z]$ son aumentados con las variables latentes $[Y, \Omega]$ en el análisis posterior.

En la estimación bayesiana se obtienen muestras de la distribución a posteriori $[\alpha, \theta, \Omega, Y|X, Z]$ a través del muestreo de Gibbs, que iterativamente simula α, θ, Ω y Y desde las distribuciones condicionales completas. Para implementar el muestreo de Gibbs, se comienza con los valores iniciales $(\alpha^0, \theta^0, \Omega^0, Y^0)$, luego se simula $(\alpha^1, \theta^1, \Omega^1, Y^1)$ y las subsiguientes mediante el mismo procedimiento descrito en la sección anterior. En la iteración j -ésima, con los valores $(\alpha^j, \theta^j, \Omega^j, Y^j)$:

- Generar Ω^{j+1} de $p(\Omega|\theta^{(j)}, \alpha^{(j)}, Y^{(j)}, X, Z)$
- Generar θ^{j+1} de $p(\theta|\Omega^{(j+1)}, \alpha^{(j)}, Y^{(j)}, X, Z)$
- Generar $(\alpha^{j+1}, Y^{(j+1)})$ de $p(\alpha, Y|\theta^{j+1}, \Omega^{j+1}, X, Z)$

Ha sido demostrado por Geman & Geman (1984) y Geyer (1992) que para un número suficientemente grande de iteraciones j , la distribución conjunta de $(\alpha^j, \theta^j, \Omega^j, Y^j)$ converge a una tasa exponencial a la distribución posterior conjunta deseada $[\alpha, \theta, \Omega, Y|X, Z]$.

2.4.4. Integración de los modelos de ecuaciones estructurales multinivel con estimación Bayesiana

En la investigación empírica, y particularmente en las ciencias sociales, es poco frecuente encontrar datos que provengan de una muestra aleatoria y de una única población. Por el contrario, los datos usualmente exhiben al menos dos posibles fuentes de heterogeneidad.

La primera son los datos de mezcla, que involucra observaciones independientes que provienen de una de las K poblaciones con distintas distribuciones, donde no hay información acerca de a cuál de las K poblaciones pertenece una observación particular.

El segundo tipo de datos heterogéneos proviene de la extracción de observaciones de distintos grupos (también llamados conglomerados o clústeres) con una estructura jerárquica conocida. Los datos analizados en esta investigación son de este tipo, donde se tienen estudiantes agrupados en centros educativos.

Como los individuos dentro de un grupo comparten ciertos factores comunes, las observaciones aleatorias están correlacionadas. Así, el supuesto de independencia entre datos observados es violado, e ignorar la estructura correlacionada de los datos y analizarlos como observaciones de una muestra aleatoria simple llevará a resultados erróneos. Además, es deseable establecer un modelo significativo para los niveles altos y estudiar los efectos de las variables latentes entre grupos sobre las variables latentes dentro de los grupos.

Sea u_{gi} , para $i = 1, \dots, N_G$, un conjunto de vectores aleatorios p -variados, anidados dentro de grupos $g = 1, \dots, G$. Los tamaños de muestra N_G pueden ser distintos entre grupos por lo que el set de datos se le llama “no balanceado”. En el primer nivel, y condicional a la media grupal v_g , se asume que las observaciones aleatorias en cada grupo satisfacen la siguiente ecuación de medición (Song & Lee 2012; Lee 2007):

$$u_{gi} = v_g + \Lambda_{ig}\omega_{1gi} + \epsilon_{1gi}$$

Donde Λ_{ig} es una matriz de cargas factoriales, ω_{1gi} es un vector aleatorio de factores latentes y ϵ_{1gi} es un vector aleatorio de errores de medición que es independiente de ω_{1gi} y se distribuye como $N[0, \Psi_{1g}]$ donde Ψ_{1g} es una matriz diagonal. Además, u_{gi} y u_{gj} no son independientes debido a la existencia de v_g .

Para tomar en cuenta la estructura multinivel se asume que la media grupal v_g satisface el siguiente modelo de análisis factorial:

$$v_g = \mu + \Lambda_2\omega_{2g} + \epsilon_{2g}$$

Donde μ es el vector de interceptos, Λ_2 es una matriz de cargas factoriales, ω_{2g} es un vector de variables latentes y ϵ_{2g} es un vector aleatorio de errores de medición que es independiente de ω_{2g} y se distribuye como $N[0, \Psi_2]$ donde Ψ_2 es una matriz diagonal. Además, se asume que los errores de medición en el primer y segundo nivel son independientes.

Combinando las ecuaciones anteriores, se tiene que:

$$u_{gi} = \mu + \Lambda_2\omega_{2g} + \epsilon_{2g} + \Lambda_{ig}\omega_{1gi} + \epsilon_{1gi}$$

Con el fin de evaluar las interrelaciones entre las variables latentes, los vectores latentes ω_{1gi} y ω_{2g} son subdivididos como $\omega_{1gi} = (\eta_{1gi}^T, \xi_{1gi}^T)^T$ y $\omega_{2g} = (\eta_{2g}^T, \xi_{2g}^T)^T$ donde η_{1gi} , ξ_{1gi} , η_{2g} y ξ_{2g} son vectores latentes para los niveles $j = 1, 2$. Los vectores ξ_{1gi} y ξ_{2g} se distribuyen como $N[0, \Phi_{1g}]$ y $N[0, \Phi_2]$ respectivamente.

Las siguientes ecuaciones estructurales son incorporadas en los modelos entre grupos y dentro de grupos del modelo de dos niveles que fue estimado en esta investigación:

$$\begin{aligned}\eta_{1gi} &= \Pi_{1g}\eta_{1gi} + \Gamma_{1g}\xi_{1gi} + \delta_{1gi} \\ \eta_{2g} &= \Pi_2\eta_{2g} + \Gamma_2\xi_{2g} + \delta_{2g}\end{aligned}$$

Donde Π_2 , Π_{1g} , Γ_2 , Γ_{1g} son matrices de parámetros desconocidos, δ_{1gi} es un vector de errores de medición distribuido como $N[0, \Psi_{1g\delta}]$, δ_{2g} es un vector de errores de medición distribuido como $N[0, \Psi_{2\delta}]$, y $\Psi_{1g\delta}$ y $\Psi_{2\delta}$ son matrices diagonales.

El modelo anterior puede ser estimado para variables observadas continuas y categóricas según lo expuesto en la sección anterior.

Para la estimación Bayesiana, nuevamente se recurre a la estrategia de aumentar los datos observados utilizando herramientas MCMC para simular las observaciones a posteriori. El procedimiento es el siguiente (Song & Lee, 2012):

Sea θ el vector de parámetros que contiene todos los parámetros estructurales desconocidos en $\Lambda_{1g}, \Psi_{1g}, \Pi_{1g}, \Gamma_{1g}, \Phi_{1g}, \Lambda_{1g\delta}, \mu, \Lambda_2, \Psi_2, \Pi_2, \Gamma_2, \Phi_2$ y $\Lambda_{2\delta}$, y sea α el vector de parámetros que contiene todos los umbrales desconocidos (para las variables categóricas).

Sea $X_g = (x_{g1}, \dots, x_{gN_g})$ y $X = (X_1, \dots, X_G)$ los datos continuos observados, y sea $Z_g = (z_{g1}, \dots, z_{gN_g})$ y $Z = (Z_1, \dots, Z_G)$ los datos categóricos observados. Sea $Y_g = (y_{g1}, \dots, y_{gN_g})$ y $Y = (Y_1, \dots, Y_G)$ las mediciones latentes continuas asociadas con Z_g y Z respectivamente. Los datos observados serán aumentados con Y en el análisis a posteriori. Una vez que Y esté dado, todos los datos serán continuos.

Sea $V = (v_1, \dots, v_G)$ la matriz de variables latentes entre grupos. Sea $\Omega_g = (\omega_{1g1}, \dots, \omega_{1gN_g})$, $\Omega_1 = (\omega_{11}, \dots, \omega_{1G})$ y $\Omega_2 = (\omega_{21}, \dots, \omega_{2G})$ las matrices de variables latentes en los niveles dentro de grupos y entre grupos.

En el análisis a posteriori, los datos observados (X, Z) serán aumentados con $(Y, V, \Omega_1, \Omega_2)$ y se considerará la distribución posterior conjunta $[\theta, \alpha, Y, V, \Omega_1, \Omega_2 | X, Z]$. El muestreo de Gibbs será utilizado para generar una secuencia de observaciones de esta distribución posterior conjunta. Para aplicar el muestreo de Gibbs se re muestrea de las siguientes distribuciones condicionales: $[V | \theta, \alpha, Y, \Omega_1, \Omega_2, X, Z]$, $[\Omega_1 | \theta, \alpha, Y, V, \Omega_2, X, Z]$, $[\Omega_2 | \theta, \alpha, Y, V, \Omega_1, X, Z]$, $[\alpha, Y | \theta, V, \Omega_1, \Omega_2, X, Z]$ y $[\theta | \alpha, Y, \Omega_1, \Omega_2, X, Z]$.

Un reto adicional de esta investigación dada la revisión de teoría con respecto a los SEM multinivel con estimación Bayesiana es la modelación de relaciones estructurales en el segundo nivel, ya que hasta el momento, toda la teoría y las aplicaciones de los modelos SEM Multinivel solo toman en cuenta la estructura jerárquica, sin prestar atención a las relaciones teórica que se puedan plantear en el nivel de grupos mediante un modelo estructural. El anexo 1 muestra el detalle de las distribuciones condicionales, siguiendo a (Song & Lee, 2012).

2.4.4.1. Estimación del SEM para datos multinivel con estimación Bayesiana en *Mplus*

En la actualidad hay disponible un número considerable de software comerciales o de código abierto para estimar modelos de ecuaciones estructurales multinivel. Estos incluyen el módulo gllamm de STATA (Rabe-Hesketh, Skrondal, & Pickles, 2004), Mplus (Muthén & Muthén, (1998-2015)), LISREL (Jöreskog & Sörbom, 1993), el paquete xxM de R (Metha & Neale, 2005), OpenBUGS (Thomas & OHara, 2004), entre otros.

En general, el uso de alguno de estos paquetes estadísticos depende de la pregunta de investigación que deba ser respondida y del diseño del modelo a ser estimado. Por ejemplo, para modelos de ecuaciones estructurales multinivel en dos niveles, todos estos software incluyen propiedades que

les permite realizar estimaciones por diferentes métodos. Si se tratara de tres niveles, la lista se reduce a cinco, mientras que para modelos de cuatro o más niveles, únicamente dos software ofrecen funcionalidades para estimarlos.

En cuanto a métodos de estimación, todos ofrecen la estimación por máxima verosimilitud (a excepción de los BUGS), y un número más reducido permiten estimaciones de errores estándar robustos o mínimos cuadrados ponderados. Únicamente con tres de estos software es posible estimar ML-SEM mediante el uso de MCMC, con un buen grado de flexibilidad en la especificación de las distribuciones a priori (siendo el lenguaje BUGS menos restrictivo que MPlus o MLwiN).

Dado lo anterior, y al tomar en cuenta otras características como la imputación de valores faltantes, los tipos de variable respuesta que puedan modelarse, los tipos de variables latentes que pueden incluirse en el modelo y la variedad de modelos que son susceptibles de estimación, los software Mplus y los programas compatibles con el lenguaje Bugs son los que presentan las condiciones requeridas para poder estimar modelos de ecuaciones estructurales para datos multinivel con estimación Bayesiana.

Finalmente, las facilidades que ofrece Mplus para el diseño de simulaciones *Monte Carlo* fue el factor considerado como determinante para dirigir las estimaciones involucradas en esta tesis hacia este software. El procedimiento para la estimación Bayesiana en Mplus se resume a continuación:

La estimación Bayesiana difiere de la estimación frecuentista en que los parámetros no se consideran constantes sino variables. A los parámetros se le puede asignar distribuciones a priori, correspondientes a la teoría o a estudios previos. Junto con la verosimilitud de los datos, esto da lugar a distribuciones a posteriori para los parámetros.

La estimación Bayesiana utiliza los algoritmos MCMC para crear aproximaciones a las distribuciones a posteriori al realizar de forma iterativa las iteraciones aleatorias en la cadena de MCMC. Las iteraciones iniciales en la cadena MCMC se conocen como la fase de “quemado”. En Mplus, la primera mitad de cada cadena se descarta como parte de esta fase de descarte (por default), aunque esto puede ser modificado por el usuario.

Mplus ofrece la posibilidad de seleccionar el tipo de estimador puntual Bayesiano a estimar (media, mediana o moda) siendo el valor default la mediana. Además, indicar el número de cadenas MCMC independientes a utilizar para realizar las estimaciones, con un valor default de dos cadenas.

Además, es posible seleccionar entre dos opciones para especificar el algoritmo de cadena MCMC a utilizar para generar la distribución a posteriori de los parámetros: GIBBS o Metropolis-Hastings. Además, las distribuciones que pueden ser utilizadas para establecer distribuciones a priori son: Normal, Lognormal, Uniforme, Gamma Inversa, Gamma, Wishart Inversa y Dirichlet.

Otras especificaciones incluyen semillas para la generación de datos aleatorios para valores iniciales de las cadenas, el número de iteraciones para evaluar convergencia, el espaciado entre iteraciones

a ser utilizadas para realizar las estimaciones y el número de procesadores a ser utilizadas para incrementar la velocidad computacional (Muthén & Muthén, (1998-2015)).

Finalmente, todo el código fue programado desde R y mediante el uso del paquete `MplusAutomation` fue realizada la estimación de los modelos en `Mplus` desde R (Hallquist & Wiley, 2018)¹⁰.

2.5. La discusión en el uso de las distribuciones a priori con varianzas pequeñas en el contexto de modelos de ecuaciones estructurales con estimación Bayesiana

Los modelos de ecuaciones estructurales se pueden separar en dos componentes: el modelo de medición o análisis factorial confirmatorio y el modelo estructural. En el CFA la rotación factorial no es aplicable debido a las restricciones identificadas previamente al fijar la mayor parte de las cargas cruzadas como iguales a cero. En otras palabras, la rotación no existe en el CFA porque una estructura simple es obtenida al especificar que los indicadores carguen en un solo factor latente y que las cargas cruzadas sean exactamente iguales a cero, lo que además beneficia la identificación del modelo al no requerir la estimación de parámetros que se considerarán como fijos, es decir, parámetros que no serán de libre estimación (Brown, 2006).

Debido a las bondades de la estimación Bayesiana en los SEM, recientemente se ha desarrollado un nuevo enfoque con la intención de producir un análisis que refleje de una mejor manera las teorías del investigador y sus creencias a priori. Esto se hace al utilizar sistemáticamente distribuciones a priori informativas para los parámetros que no deberían ser estimados libremente de acuerdo con la teoría del investigador.

En el análisis frecuentista tales parámetros son fijados como iguales a cero (o son fijados a ser iguales a algún otro parámetro del modelo), pero si estos parámetros fueran liberados y debieran ser estimados el modelo presentaría problemas de identificación y muy posiblemente de estimación.

Mediante el enfoque Bayesiano, por el contrario, se logra la identificación del modelo al impulsar el uso de varianzas a priori muy pequeñas para estos parámetros. Muthén y Asparouhov (2012(1)) llaman a este enfoque Bayesian Structural Equation Modeling (BSEM, siglas en inglés), que es básicamente una estimación Bayesiana de un modelo SEM sin restringir todos los parámetros que debieran ser considerados como fijos a valores predeterminados.

¹⁰ Agradezco a los autores de la librería `MplusAutomation` por haber atendido a mi solicitud de agregar dos estadísticos adicionales en la opción `"summary"` del comando `"read.Models"` de las simulaciones Monte Carlo generadas desde `MPlus`, en julio del año 2019.

En la sección 2.1 fue presentado un ejemplo de un CFA de tres variables latentes, cada una medida por tres variables indicadoras, según la siguiente matriz:

$$\Lambda_{(9 \times 3)} = \begin{pmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \\ 0 & 0 & \lambda_{73} \\ 0 & 0 & \lambda_{83} \\ 0 & 0 & \lambda_{93} \end{pmatrix}$$

Cada variable indicadora carga únicamente en un factor latente, y las cargas cruzadas han sido fijadas como cero, lo que significa que cada variable indicadora no está influenciada por otros factores distintos del principal.

Una carga factorial fijada exactamente como cero puede ser vista como una distribución a priori que tiene media cero y varianza cero. Según los autores del enfoque BSEM, para reflejar de una mejor manera las teorías de los investigadores se debe utilizar una distribución a priori con media cero y una distribución normal con una varianza muy pequeña.

Continuando con el ejemplo de la matriz anterior, donde $\lambda_{12} = 0$, se podría utilizar una distribución normal como $\lambda_{12} \sim N(0, 0.1)$ con lo que el 95% de la variación de la carga factorial estaría entre -0.2 y $+0.2$, lo que es considerado como una muy baja carga factorial, por lo que esta distribución a priori esencialmente significa que la carga cruzada es cercana a cero, pero no exactamente igual a cero.

El uso de varianzas a priori muy pequeñas para las cargas cruzadas introduce información en el análisis Bayesiano que evita los problemas de identificación que se presentarían si se aplicara en el enfoque frecuentista (Muthén & Asparouhov, 2012 (1)). La escogencia de la varianza queda a criterio del investigador y la teoría. Además, la varianza a priori se debe determinar en relación con la escala de las variables observadas y latentes.

De acuerdo con los autores (ídem), una ventaja adicional del BSEM es que produce distribuciones a posteriori para las cargas cruzadas que pueden ser utilizadas para los índices de modificación tradicionalmente estimados por varios software estadísticos.

En Muthén & Asparouhov (2012(1)), tomando como base el análisis realizado por Holzinger y Swineford (1939), de un análisis factorial confirmatorio de cuatro dominios o factores latentes y 19 variables indicadoras en el contexto de medición de habilidades de estudiantes de séptimo y octavo grado de Estados Unidos, donde cada variable indicadora carga en uno y solo un factor latente (entre 4 y 6 variables por constructo), los autores ejemplifican la aplicación práctica de la propuesta BSEM.

Muestran que con los criterios de evaluación tradicionales bajo la estimación frecuentista como el Chi-cuadrado, el RMSEA y el CFI, el modelo factorial confirmatorio es rechazado según los estándares de aceptación de estos índices de evaluación de bondad de ajuste global, y de forma inversa, cuando se realizan análisis factoriales exploratorios del mismo modelo (donde todas las variables indicadoras pueden cargar libremente en todos los factores latentes), el ajuste del modelo mejora, y se mantienen las principales cargas factoriales en las variables hipotetizadas en el CFA; sin embargo, con el EFA varias de las cargas cruzadas pasan a tener efectos significativos.

Al realizar la misma estimación del modelo pero con el enfoque Bayesiano (fijando las cargas factoriales cruzadas como iguales a cero), el modelo continúa sin mostrar un buen nivel de ajuste medido por la probabilidad asociada a la prueba cociente de verosimilitudes del modelo nulo respecto del modelo irrestricto (utilizando el estadístico *P-value predictivo a posteriori*), lo que está en línea con el ajuste observado en la estimación frecuentista.

Al aplicar la metodología descrita en el enfoque BSEM (estimación Bayesiana con cargas factoriales cruzadas siguiendo una distribución a priori normal con media 0 y varianza 0.01), el ajuste del modelo se considera aceptable (los límites la probabilidad asociada a la prueba cociente de verosimilitudes del modelo nulo respecto del modelo irrestricto entre el 2.5% y el 97.5% incluyen el cero)¹¹.

Adicional a este ejemplo, en el documento los autores muestran la aplicación del enfoque BSEM en otros parámetros que son tradicionalmente fijados en cero en el contexto del CFA, como las correlaciones residuales y las restricciones en los coeficientes estructurales del modelo estructural.

Después de su publicación, el documento fue analizado y debatido por otros autores. En el año 2012, MacCallum y Edwards publicaron un comentario titulado *Esperanzas y precauciones en la implementación del enfoque BSEM*. En este, los autores concuerdan en los potenciales beneficios de este enfoque BSEM, destacando la especificación de modelos más realistas, y plantean algunas posibles limitaciones técnicas en la implementación de este enfoque.

En cuanto a la especificación de un modelo convencional que incluya distribuciones a priori con varianzas pequeñas en los parámetros usualmente fijados a cero, plantean que el investigador enfrenta dos decisiones que pueden afectar los resultados. La primera, la escogencia de los parámetros que serán estimados con el nuevo enfoque, y la segunda, en cuanto a la definición de *varianza pequeña* para la estimación de esos parámetros.

Sobre esto último, mencionan que la escogencia de distribuciones a priori puede tener implicaciones para la especificación del modelo que no son deseables y, a su vez, puede afectar la computación del algoritmo MCMC así como sus resultados, por lo cual recomiendan a los investigadores que apliquen el enfoque BSEM a realizar análisis de sensibilidad en la especificación de las distribuciones a priori para esos parámetros (MacCallum & Edwards, 2012, pág. 3).

¹¹ En la sección del diseño de las simulaciones serán explicados en mayor detalle las propuestas para evaluar y comparar ajustes de modelos.

En materia de los asuntos computacionales del algoritmo MCMC, afirman que el enfoque BSEM complica la garantía de la identificación del modelo, ya que la propuesta no sigue la línea tradicional de aplicar algunas reglas simples para garantizar la identificación del modelo a ser estimado, lo cual dificulta determinar a partir de qué punto el modelo se vuelve no identificado, ya que entre mayor sea la varianza de la distribución a priori, más se acerca el modelo al punto de no identificación¹² (MacCallum & Edwards, 2012, pág. 4).

En otra revisión al enfoque BSEM, Rindskopf (2012) muestra extensiones adicionales y adaptaciones de los métodos propuestos por Muthén & Asparouhov (2012(1)) para evidenciar que los investigadores no Bayesianos pueden tomar algunas de sus ventajas al utilizar restricciones de intervalos en los parámetros, que permitan a los investigadores frecuentistas utilizar software para la modelación de ecuaciones estructurales estándar de una forma similar a lo que la estimación Bayesiana produce al utilizar distribuciones a priori sobre esos parámetros.

En respuesta a ambos autores, Muthén & Asparouhov (2012(2)) publican una réplica abordando los retos o cuestionamientos presentados anteriormente. Sobre el posible mal uso al enfoque BSEM en la escogencia de los parámetros, Muthén & Asparouhov establecen que dicha escogencia dependerá del estado de la evidencia que exista en torno a cada investigación particular. En etapas tempranas de una investigación poco abordada en la literatura, donde el foco es la exploración, las distribuciones a priori debieran ser no informativas o débilmente informativas (grandes varianzas), permitiendo a los datos ser el factor que más influirá en la estimación, pero conforme la evidencia empírica crezca a través de nuevos estudios, podrían ser introducidas más distribuciones a priori informativas en línea con el enfoque BSEM (Muthén & Asparouhov, 2012 (2), pág. 347).

Con respecto a la ganancia en el ajuste del modelo, Muthén & Asparouhov (2012 (2))) indican que el hecho de que el modelo BSEM muestre mejor ajuste cuando dichas estimaciones de parámetros se desvíen sustancialmente de cero puede provocar un mal uso del método. Estas desviaciones pueden ocultar importantes especificaciones erróneas del modelo, como la necesidad de un número diferente de factores o diferentes relaciones estructurales. Muthén & Asparouhov (2012 (2), pág. 347) enfatizan que si se observan desviaciones significativas y sustancialmente alejadas del cero, es importante investigar las razones de ese hallazgo.

MacCallum et al. (2012) señalaron que cuando la cadena de Markov Monte Carlo se aplica a modelos de variables latentes, los signos de carga factorial para un factor dado pueden cambiar durante las iteraciones de MCMC (entre positivos y negativos). Muthén & Asparouhov (2012) afirman que con la estimación de máxima verosimilitud se sabe que el cambio de signo corresponde a una inversión ignorable de la dirección del factor. Sin embargo, en cadenas MCMC, esto puede llevar a obtener resultados sin sentido que promedian las cargas positivas y negativas. Para los autores, no está claro qué tan común es este fenómeno para los modelos de variables latentes, pero en su propia

¹² Un modelo está sub-identificado cuando el número de parámetros desconocidos (estimados libremente) excede el número de piezas de información conocida (el número de elementos de la matriz de varianzas y covarianzas). Un modelo sub-identificado no puede ser resuelto, porque hay un infinito número de parámetros estimados que resultan en ajuste perfecto del modelo.

experiencia consideran que no es común en modelos con grandes cargas factoriales y errores estándar pequeños (Muthén & Asparouhov, 2012 (2), pág. 347).

Muthén & Asparouhov adicionan además que aunque los cambios de signo pueden ocurrir, estos pueden detectarse en los gráficos producidos en el análisis Bayesiano del software de su autoría, *Mplus*. Además, para evitar el cambio de signo en los modelos factoriales, introdujeron una corrección en el algoritmo de reetiquetado que, para cada iteración de MCMC y para cada factor, calcula la suma de todas las cargas factoriales y cambia los signos de todas las cargas si esta suma es negativa (Asparouhov y Muthén, 2012), lo que induce una inversión de la dirección del factor.

En años más recientes, la discusión ha girado hacia el uso del enfoque BSEM en el estudio de la invariancia factorial entre muchos grupos pertenecientes a una misma población. Muthén y Asparouhov (2013(1)) generalizan el uso del BSEM al análisis de la medición de la invariancia a través de varios grupos o varios momentos en el tiempo, aplicando la idea de una distribución a priori con media cero y varianza pequeña a los parámetros de medición, por lo que se refieren a esto como la invariancia de medición aproximada.

El enfoque BSEM es aumentado con la estimación de las diferencias entre cada parámetro de medición y su promedio a través de todos los grupos. Asumiendo que el parámetro es aproximadamente invariante para la mayoría de grupos, estas diferencias pueden apuntar a los grupos que tienen una no-invariancia significativa. Posterior a este paso, la restricción de igualdad de los parámetros que fueron identificados como no-invariantes para los grupos puede relajarse y volverse a realizar la estimación Bayesiana únicamente manteniendo iguales entre los grupos los parámetros invariantes (Muthén & Asparouhov, 2013(1)).

Una extensión del estudio de invariancia presenta en Muthén & Asparouhov (2012(2)), donde los autores describen dos modelos que se diferencian entre sí en la forma de estimación de la media y varianza del factor, y los supuestos que son agregados a la información de los datos.

Posteriormente a la presentación del estudio de la invariancia, en un número especial del *Journal of Management* dedicado a la estimación Bayesiana (Zyphur & Oswald, 2015), Stromeier, Miller, Sriramachandramurthy y DeMartino (2015) retoman la discusión del enfoque BSEM. Los autores realizan una serie de críticas y recomendaciones. Con respecto a la posibilidad de obtener cargas factoriales significativas que el investigador planteó como cercanas a cero, argumentan que esto puede llevar a la sub estimación de correlaciones entre factores. Sin embargo, en la respuesta a esta crítica, Asparouhov, Muthén y Morin (2015) muestran, mediante estudios simulados, como su crítica no tiene sustento empírico.

En la crítica a las covarianzas residuales y su estimación con el enfoque BSEM, Stroyemer y otros (2015) recomiendan evitar esa aplicación y concluyen que no debe utilizarse a la ligera hasta que hayan publicados más estudios con esta línea de investigación. Asparouhov y otros (2015) muestran también, mediante estudios de simulación, que la aplicación puede realizarse siempre y cuando el investigador posea evidencia para plantear esos patrones de relaciones, y no sólo como un método para mejorar el ajuste global del modelo.

Si bien es reciente la discusión sobre el uso de varianzas pequeñas en las distribuciones a priori de parámetros que usualmente son fijados a ser cero en el contexto de las ecuaciones estructurales, nada ha sido discutido acerca de las implicaciones de su aplicación en el contexto de ecuaciones estructurales multinivel. Es por esta razón que esta investigación pretende aportar a la literatura y a esta discusión al evaluar el desempeño de modelos de ecuaciones estructurales para datos multinivel con estimación Bayesiana aplicando el enfoque BSEM, lo que en esta tesis será referido como ML-BSEM.

El enfoque BSEM es aplicable a todos los parámetros restringidos de un modelo SEM, como las covarianzas residuales o los parámetros estructurales, pero en esta investigación únicamente fueron aplicados a las cargas factoriales del modelo de medición, como será explicado en la sección 4 de esta investigación.

3. Estado de la cuestión sobre el caso aplicado

El principal aporte que realiza un modelo de ecuaciones estructurales es que permite a los investigadores evaluar modelos teóricos e hipótesis de causalidad, siendo una de las herramientas más potentes para el estudio de relaciones causales sobre datos no experimentales. Aunque teóricamente no prueban la causalidad en un sentido estricto, los modelos de ecuaciones estructurales son una de las pocas técnicas estadísticas multivariadas que pueden dar evidencia de soporte sobre una hipótesis de causalidad¹³.

A diferencia de los análisis estadísticos exploratorios, los modelos de ecuaciones estructurales se centran en probar hipótesis en modelos con ciertas restricciones sobre los parámetros previamente identificadas. El objetivo es probar si el modelo con estas restricciones ajusta bien a los datos con los cuáles se buscará realizar la prueba.

De esta manera, los modelos de ecuaciones estructurales no pueden ser utilizados con fines exploratorios, y por el contrario, únicamente son útiles cuando se intenta probar hipótesis previamente establecidas.

Para poder establecer hipótesis es necesaria una justificación teórica de la misma, basada en la literatura sobre el tema analizado y de la guía de expertos en la materia analizada. Por tal razón, en esta sección se realiza una revisión de la literatura referida a los factores asociados al rendimiento educativo, los tipos de análisis estadísticos que se han realizado en relación con esto, y las posibles relaciones entre las variables predictoras.

3.1. Breve descripción de los factores asociados al rendimiento educativo

Usualmente, las políticas educativas se centran en proveer factores fácilmente cuantificables en los sistemas educativos, como dinero, infraestructura, equipo y mobiliario, libros, entre otros. Estas políticas son las más usadas porque son fáciles de administrar y son visibles a la sociedad en general, por lo que son políticamente viables. Pero mejorar o incrementar la dotación de estos factores no se traduce necesariamente en una mejora en el proceso de aprendizaje de los estudiantes.

Empíricamente, identificar el grado en que diferentes variables contribuyen al aprendizaje de los estudiantes es tarea difícil por múltiples razones. Estos factores son numerosos y complejos, y pueden afectar a los estudiantes de diferentes maneras dependiendo de su etnia, estrato socioeconómico, género u otras características. Además, pueden interactuar con otros factores y producir resultados inesperados.

El impacto de los recursos en el aprendizaje de los estudiantes puede ser limitado porque estos no necesariamente son colocados con el propósito de mejorar el aprendizaje del estudiante. Los centros educativos y los sistemas escolares son lugares altamente politizados donde las decisiones

¹³ Algunas otras son las redes bayesianas y las redes neuronales.

son tomadas por un número de razones, de las cuales el mejoramiento del aprendizaje de los estudiantes puede ser solo una. (Vegas & Petrow, 2007)

Así, los estudios que analizan el rendimiento educativo normalmente clasifican a todos los factores que podrían afectar el proceso de aprendizaje de un estudiante dentro de tres categorías generales: a) Factores del estudiante, b) Factores escolares y c) Factores institucionales y de política educativa. Se procede a enumerar algunos de ellos siguiendo lo recopilado por Vegas & Petrow (2007).

3.1.1. Factores del Estudiante

Estos son los factores que el estudiante trae consigo antes de ingresar al sistema educativo, tanto características propias del estudiante como del entorno familiar. La edad a la que el estudiante ingresa a la educación primaria y la preparación recibida antes de ingresar a la primaria son dos de los factores que han tomado relevancia en recientes investigaciones (véase Urzúa & Veramendi, 2011).

Otros factores, como la salud, la zona de residencia, el lenguaje o la etnia son importantes a la hora de analizar las diferencias en rendimiento educativo: niños mal nutridos obtienen peores rendimientos o poblaciones indígenas o rurales en general obtienen resultados inferiores a los estudiantes de zonas urbanas.

No obstante, son los factores familiares y el apoyo que el niño recibe en su hogar los que usualmente tienen los mayores efectos en la explicación del rendimiento cognitivo. El estatus socioeconómico de la familia del estudiante y el ingreso familiar han sido las variables más utilizadas como proxy a estos factores. Algunos estudios utilizan otras variables tales como el nivel de escolaridad de la madre, el estatus ocupacional del padre y los recursos educativos presentes en el hogar.

Sin embargo, estas variables por sí solas no pueden dar cuenta de la dinámica no observable del hogar, y que a menudo oculta lo que realmente está pasando en el mismo, tales como la participación de los padres en las acciones relativas de sus hijos, y la motivación para apoyar la educación de sus hijos. Por esto, resulta importante no subestimar la influencia de las acciones de los padres en el hogar sobre el rendimiento de sus hijos. (Urzúa & Veramendi, 2011).

3.1.2. Factores Escolares

Estos son las dotaciones y los recursos que ofrecen los centros educativos y que terminan afectando el nivel de rendimiento de un estudiante. Pueden clasificarse en dos categorías: las características del centro educativo y las características de los profesores.

En el caso de las características de los centros educativos, las variables más utilizadas para capturar su efecto son la cantidad de libros que ofrecen los centros educativos, las bibliotecas, y otras

facilidades de la infraestructura propia del centro educativo, como salas de estudio para los estudiantes, tamaño y comodidades del aula, entre otros.

Otra de las variables que ha tomado relevancia en estudios recientes es el acceso y aprendizaje de las tecnologías de información y comunicación (TICs), ya que estas están cada vez más presentes en la vida cotidiana y en el trabajo.

Así, muchos países han optado por políticas educativas en la que se destinan una mayor cantidad de recursos económicos para instalar o incrementar el equipamiento tecnológico de los centros educativos. Invertir en TICs es visto en el plano educativo como un camino primario para preparar a las personas a las nuevas demandas del mercado laboral y, a la vez permitir que las poblaciones más vulnerables puedan desarrollar las competencias necesarias para tener igualdad de acceso a oportunidades de empleo. (Urzúa & Veramendi, 2011)

Por otra parte, los profesores pueden tener un impacto significativo en el rendimiento de los estudiantes, ya que estos son los que participan directa y continuamente en el proceso de aprendizaje de los mismos, y por ende, un profesor poco calificado o con métodos de enseñanza inefectivos puede afectar negativamente el rendimiento de los estudiantes, y viceversa.

De acuerdo con Vegas & Petrow (2007) cada vez más evidencia apoya la noción intuitiva de que los profesores juegan un rol clave en qué, cómo y cuánto aprenden los estudiantes. Atraer individuos calificados a la profesión de la enseñanza, retener a los profesores calificados, proveerlos de las habilidades y conocimientos necesarios y motivarlos para realizar el mejor trabajo posible es uno de los principales retos de los sistemas educativos.

Algunas características del cuerpo docente, como su formación profesional, años de experiencia o cursos de capacitación específicos son fáciles de medir y utilizar para el análisis de la influencia de los profesores en los estudiantes, sin embargo, algunas otras como el nivel de motivación, sus métodos de enseñanza, la actitud hacia la materia y hacia los estudiantes son difíciles de medir, especialmente en sistemas educativos donde los profesores no son evaluados rigurosamente para medir su desempeño y donde su continuidad en el personal docente de un centro educativo no depende de su desempeño.

3.1.3. Factores institucionales y de política educativa

La estructura organizacional de un sistema educativo juega un papel importante en el aprendizaje de los estudiantes. La manera en que estén distribuidas las responsabilidades de los tomadores de decisiones en temas como financiamiento, gasto, movilidad del personal (contratar y despedir profesores), independencia para establecer las metodologías de enseñanza, entre otros, son factores institucionales que pueden afectar el rendimiento de los alumnos.

3.2. Estudios sobre factores asociados al rendimiento educativo

La literatura sobre factores asociados con el rendimiento educativo o cognitivo es vasta. El Informe Coleman (Coleman, 1966) fue uno de los primeros estudios desarrollado sobre este tema y aún tiene una influencia significativa en las investigaciones acerca del rendimiento educativo. En él se sugiere que las diferencias en los factores escolares tienen poco que ver con las diferencias en rendimiento, mientras que los factores familiares son más importantes.

En esta misma línea, Hanushek (1986) realiza una revisión de todos los estudios que habían sido realizados hasta mediados de la década de 1980, encontrando que la evidencia del efecto del nivel de gasto por estudiante u otros factores escolares en el rendimiento educativo es extremadamente débil y desaparece cuando se toman en cuenta las diferencias en los factores familiares.

Desde entonces se han realizado una gran cantidad de investigaciones sobre este tema, utilizando distintas bases de datos en decenas de países, con distintas metodologías estadísticas, para varios niveles educativos y en general, estudiando diversas aristas de los factores que podrían tener efectos en la mejora del rendimiento educativo.

Al ser tantos los estudios no es posible analizarlos todos y cada uno en esta sección. Sin embargo, resulta de vital importancia el poder garantizar el análisis de al menos algunos de los más representativos, especialmente por la rigurosidad de la metodología estadística aplicada.

En un artículo titulado *School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010*, Glewwe y otros (2011) realizan una revisión de las investigaciones sobre factores asociados al rendimiento educativo, pero limitando el periodo de análisis desde el año 1990 y hasta el año 2010, justificando el utilizar únicamente este periodo y no años anteriores en que, a su criterio, los estudios más antiguos en esta materia se encontraban muy limitados en el tipo y calidad de las metodologías estadísticas que podían ser aplicadas, y que por lo tanto en los estudios más recientes pudieron aplicarse análisis estadísticos más sofisticados.

Para esto, Glewwe y otros (2011) realizaron un estudio de meta-análisis de trabajos que analizaran los factores asociados al rendimiento educativo. Su revisión de literatura se enfocó en investigaciones sobre educación primaria y secundaria, es decir, no incluye educación preescolar, vocacional o universitaria, y que utilizaran el aprendizaje escolar como variable respuesta (usualmente medido en términos de puntajes en pruebas).

Además, debido al consenso de que las variables personales y familiares son en general las que muestran más relación con resultados académicos, y a la ambigüedad de los efectos de variables de los centros educativos y de los profesores, limitan su enfoque al análisis de investigaciones que utilizan variables escolares y del cuerpo docente, sin entrar en consideración de los factores personales y familiares.

Para el meta-análisis, la estrategia utilizada fue buscar una gran variedad de fuentes (más de 9 mil artículos científicos) y eliminar sistemáticamente los trabajos que no cumplieran con una serie de criterios definidos por Glewwe y otros (2011) acerca de relevancia y calidad, lo que los lleva a

analizar 79 artículos científicos, de los cuales 43 son clasificados como de “alta calidad” por las metodologías estadísticas aplicadas¹⁴.

Aprovechando esta revisión sistemática de literatura, esta sección se basa en los resultados de los 43 artículos más relevantes de acuerdo con los criterios de los autores.

3.2.1. Diseños experimentales

De los 43 estudios seleccionados, 13 corresponden a diseños experimentales. A diferencia de los estudios que utilizan datos observacionales, los diseños experimentales sí permiten probar hipótesis de causalidad planteadas previamente por los investigadores, asignando aleatoriamente a los individuos en los grupos de control y de tratamiento, y evaluando el efecto de un tratamiento al comparar ambos grupos.

En el área de educación el uso de diseños experimentales no es tan común debido a problemas metodológicos propios que surgen de la compleja interacción entre individuos que no permite controlar todos los factores dentro del experimento, así como por cuestiones éticas y políticas.

Sin embargo ha habido esfuerzos por desarrollarlos para analizar factores asociados al rendimiento educativo, aunque la mayoría se enfocan en el análisis de programas específicos que podrían tener un impacto sobre el desempeño escolar y que pueden ser manipulados fácilmente en un experimento (Duflo & Banerjee, 2017), por lo que no fue posible encontrar en la literatura a la que se tuvo acceso, ensayos aleatorios controlados que estudiaran las características de los profesores o de los directores, debido a que estos son muy difíciles de aleatorizar.

Las características de infraestructura y de materiales pedagógicos que fueron analizados son los libros de texto, las computadoras y las pizarras. Dos de los experimentos examinan los libros de texto, uno en Filipinas (Tan, Lane, and Lassibille, 1999) y otro en Kenya (Glewwe, Kremer, and Moulin, 2009). Los resultados en ambos sugieren que no hay un impacto positivo de proveer libros de texto.

La variable que ha sido más estudiada (en 5 de los 13 experimentos) es la disponibilidad de computadoras y aparatos electrónicos para uso pedagógico. La evidencia a la fecha del impacto de

¹⁴ Esta búsqueda inicial llevó a los autores a tener más de 9 mil artículos que abarcaban el tema. Luego, dos de los autores revisaron cada uno de los más de 9 mil artículos individualmente, seleccionando aquellos que parecían potencialmente relevantes basados en la información encontrada en el resumen (abstract) y en algunos casos en la introducción o conclusión de los artículos. Los artículos que no cumplían con los requerimientos de enfocarse en países desarrollados o no estimaban el efecto de variables de los centros educativos o de los docentes sobre el rendimiento de los estudiantes fueron eliminados. Este proceso redujo el número de artículos a 253. Estos 253 fueron efectivamente leídos por completo, y mediante el criterio de los autores sobre su relevancia, impacto y sofisticación de la metodología estadística fueron retenidos sólo 79 artículos. Finalmente, estos 79 fueron separados en dos categorías, donde se analizan por aparte 43 que tienen una “alta calidad” de acuerdo con los autores debido a que utilizan técnicas estadísticas más apropiadas para controlar el efecto de variables inobservables como diseños experimentales, modelos de diferencias en diferencias, diseños de regresiones discontinuas o métodos de emparejamiento (matching) en contraste con el uso únicamente de técnicas como regresiones lineales.

intervenciones que simplemente proveen dispositivos computacionales sugieren impactos nulos o negativos en pruebas estandarizadas (Barrera-Osorio and Linden 2009; Cristia et al. 2010; Beuermann et al. 2015; Malamud and Pop-Eleches 2011). Por otra parte, las intervenciones que se centran en el uso de tecnología para mejorar la práctica pedagógica típicamente encuentran más resultados positivos (Duflo & Banerjee, 2017).

Por ejemplo, en su artículo Banerjee, Cole, Duflo, and Linden (2007) evalúan una intervención en escuelas primarias de la India en la que los profesores recibieron capacitación en el uso de software para la enseñanza de la matemática dentro de la clase. En las escuelas pertenecientes al grupo de tratamiento, los estudiantes usan el software de dos a cuatro horas a la semana. Después de dos años de tratamiento, se encontró evidencia de que los estudiantes del grupo de tratamiento tuvieron puntuaciones significativamente más altas en pruebas matemáticas que el grupo de control, pero no se hallaron diferencias significativas en los puntajes en las pruebas de lectura.

Otro estudio (Barrera-Osorio and Linden, 2009) evaluaron el programa de Computadoras para la Educación en Colombia y encontraron resultados menos concluyentes. En este programa, los profesores recibieron computadoras y ocho meses de capacitación en su uso en la clase, y al final del experimento no se encontraron diferencias en los resultados de las pruebas matemáticas y de lectura evaluadas entre el grupo de control y tratamiento al combinar los resultados de los grados escolares analizados (de tercero a noveno), e incluso resultados positivos en noveno grado pero resultados significativamente negativos en octavo grado.

En otro estudio, Cristia y otros (2010) evalúan la hipótesis de si el incremento en el acceso a las tecnologías de información y comunicación puede aumentar los años de educación completos en Perú, ya que en el 2004 se implementó un programa educativo financiado por el Banco Interamericano de Desarrollo en el que se distribuyeron 10 computadoras en 350 centros educativos públicos de secundaria; concluyen que no hay evidencia de que la política tuviera un impacto significativo sobre los rendimientos educativos.

En un estudio similar realizado también en Perú, Cristia (2012) evalúan en un experimento aleatorio otro programa que ha tenido gran auge a nivel mundial llamado Una Computadora por Estudiante (*One Laptop Per Child Program*, OLPC). Este programa busca mejorar el aprendizaje en las regiones más pobres del mundo proveyendo computadoras personales a los niños para utilizar tanto en la escuela como en el hogar. Perú es el país en el que este programa se ha implementado a mayor escala, incluyendo la distribución de 902.000 computadoras en 319 escuelas públicas en comunidades pobres y rurales. Llegan a la siguiente conclusión: *“Los resultados sugieren efectos limitados sobre el rendimiento escolar de los niños, pero impactos positivos en las habilidades cognitivas y competencias relacionadas con el uso de las computadoras”*.

Además del uso de computadoras, otros experimentos han analizado la forma en que los centros educativos se encuentran organizados. Por ejemplo, Muralidharan y Sundararaman (2008) estudian el impacto del tamaño de clase en el rendimiento educativo en la India.

En el experimento aleatorio controlado, el tamaño de clase fue intencionalmente reducido en el grupo de tratamiento mediante la contratación de más docentes para distribuir a los estudiantes. El artículo muestra cinco estimaciones del impacto del tamaño de la clase en rendimiento de los estudiantes, de las cuales tres son negativas (como es esperado, a mayor tamaño de clase menor rendimiento) y estadísticamente significativas mientras que las otras dos estimaciones no resultaron significativas.

Sin embargo, Glewwe y otros (2011) sugieren que este experimento no puede tomarse como válido debido a que los autores no lograron separar el efecto del tamaño de clase del efecto de la contratación de los nuevos docentes; además, al ser el único experimento que hallaron que analiza este tema consideran que es difícil poder generalizar estos resultados.

Otro de los experimentos consideró el impacto de proveer alimentos en los centros educativos (Tan, Lane, and Lassibille, 1999) encontrando en resultado no significativo de este tipo de programas en Filipinas. Las clases remediales impartidas por tutores particulares también han sido estudiadas en el Programa de tutorías Balsakhi en la India (Banerjee, Cole, Duflo, and Linden, 2007), en el que los autores encontraron que el proveer de estos tutores a niños que se encuentran rezagados en el currículo académico aumenta significativamente sus puntajes en las pruebas.

3.2.2. Otros estudios internacionales

En contraste con los diseños experimentales, hay centenas de estudios que utilizan datos observacionales para realizar sus análisis. Al analizar los que aplican metodologías estadísticas más avanzadas que la típica regresión lineal, Glewwe y otros (2011) hallan que en 30 estudios, la premisa de que los libros de texto y materiales similares mejoren el aprendizaje de los estudiantes es bastante débil, por lo que no es posible tener claridad sobre su efecto.

Por el contrario, en la mayoría de estas investigaciones que utilizaron variables como la calidad y cantidad de escritorios, mesas y sillas dentro del aula encuentran una relación positiva entre estos y el rendimiento educativo, mientras que los que utilizan variables como computadoras, conexión a internet y otras relacionadas encuentran resultados débiles de asociación con la variable respuesta. Estos hallazgos en conjunto con los presentados en la subsección de experimentos muestran que en el tema de las tecnologías de información y comunicación los resultados aún no son concluyentes, y deben tomarse con precaución al momento de establecer políticas educativas.

Al estudiar las variables relacionadas con los docentes, dos de las variables más utilizadas son el nivel educativo de los profesores y su experiencia. En su revisión, Glewwe y otros (2011) encuentran resultados ambiguos, ya que en trece de las investigaciones, diez estiman resultados no significativos (e incluso divididos entre efectos positivos y efectos negativos), y mientras que dos encuentran resultados positivos significativos, un estudio muestra un resultado estadísticamente significativo pero negativo, al contrario de lo que la teoría muestra que debería ser la relación entre estas variables y el aprendizaje de los estudiantes.

Otras variables como la competencia de los profesores y su conocimiento de la materia que imparten sí muestran efectos positivos muy fuertes. De las 20 estimaciones de este tipo en los estudios seleccionados por el meta-análisis (Glewwe y otros, 2011), todas estiman efectos positivos y en 13 de estas se reportan efectos estadísticamente significativos. Estos resultados muestran bastante evidencia en favor de la hipótesis de que el conocimiento real de los profesores (y no su aproximación mediante su titulación académica) juega un rol sumamente importante en el aprendizaje de los estudiantes.

En el caso de las variables relacionadas con el centro educativo, una de las más utilizadas es la proporción de estudiantes por profesor. En la mayoría de los estudios que incorporan esta variable en sus modelos el efecto estimado es negativo, como lo explica la teoría, pero no en todos la estimación es significativa, lo que nuevamente sugiere que si bien disminuir el tamaño de una clase tiene un impacto positivo en el aprendizaje de los estudiantes, su efecto podría no ser tan contundente como se esperaría.

Al estudiar el abstencionismo de los profesores, los seis estudios que lo hicieron encontraron todos efectos negativos y en cuatro de estos el efecto es significativo. Otras variables son el número de horas de clases diarias y las tutorías, con resultados ambiguos, ya que algunos estudios encuentran efectos positivos, otros negativos, y la mayor parte de estos con poca significancia estadística.

3.2.3. Revisión de literatura de Costa Rica

De la revisión de literatura anterior, es evidente que no hay un claro consenso sobre el rol de los factores escolares y del personal docente sobre el rendimiento educativo de los estudiantes y, en general, tampoco hay evidencia para afirmar que lo que funciona para un país pueda funcionar para otro.

Por esta razón, además del estado de la cuestión que se puede obtener de la literatura internacional, resulta importante conocer lo que se ha estudiado en Costa Rica en este tema, con la finalidad de tener un mejor panorama del efecto de algunas variables en el caso del país.

La primera estimación del efecto de diversos factores en el rendimiento educativo en Costa Rica data de 1980 (Díaz y Jiménez, 1980). Mediante la estimación de una función de producción educativa con datos a nivel cantonal, los autores concluyen que en el rendimiento educativo influyen tanto los insumos escolares como los factores socioeconómicos. Si bien el estudio presenta una serie de limitaciones especialmente en la limitación de los datos disponibles, es una primera aproximación a la relación entre insumos escolares y socioeconómicos en el rendimiento educativo.

Moreira (2009) analiza la influencia de diversos factores en las puntuaciones obtenidas por estudiantes de undécimo año de colegios académicos diurnos en la prueba nacional de bachillerato de matemáticas del año 2004; a través del análisis multinivel, el autor deduce que los factores endógenos como el historial académico del estudiante, y específicamente la condición de repitencia y el nivel educativo de los padres, se relacionan con diferencias en las puntuaciones obtenidas. Los resultados se refieren a una muestra probabilística, estratificada y proporcional de estudiantes

provenientes de colegios públicos académicos diurnos de Costa Rica. Se utilizó un modelo de regresión múltiple.

Rojas (2004) estudia mediante un modelo jerárquico los factores asociados a la repitencia de los alumnos en séptimo año de colegios académicos, diurnos y públicos, con una muestra de 1442 estudiantes y 115 docentes que impartían clases de las asignaturas de Inglés, Español, Estudios Sociales, Ciencias y Matemáticas. Encuentra que variables como una mala relación de los padres con el estudiante así como una mala interacción entre profesor y estudiante tienen una relación directa con la repitencia

Además. Muestra que factores como una mayor autoestima, una más alta motivación, realizar mayor cantidad de tareas, recibir ayuda para realizar esas tareas, respeto percibido por los estudiantes de sus iguales, residir con ambos padres durante el tiempo lectivo, indicadores de condición socioeconómica del estudiante y su familia, e indicadores de la infraestructura e instalaciones del centro educativo, incluyendo el acceso a internet son factores que tienen una relación inversa con la condición de repitencia.

Por otra parte, desde el año 2009 el Ministerio de Educación Pública (MEP), específicamente el Departamento de Evaluación Académica y Certificación, de la Dirección de Gestión y Evaluación de la Calidad, ha venido realizando estudios de factores asociados al rendimiento para sus pruebas de diagnóstico de sexto y noveno grados con modelos de regresión múltiple y multinivel. Se utilizan muestras estratificadas aleatorias de conglomerados completos, donde el conglomerado es el centro educativo. Se definen tres niveles de agregación: estudiante, docente y director. Los participantes de cada nivel deben completar un cuestionario de contexto para luego relacionar esas variables con el rendimiento de los estudiantes en las pruebas. El primer estudio de este tipo se realizó para identificar predictores en las pruebas diagnósticas de sexto grado aplicadas en el año 2008 (MEP, 2010).

Más recientemente el MEP ha publicado el estudio de factores asociados al rendimiento en las pruebas diagnósticas de noveno (MEP, 2012). Uno de los méritos que poseen estos estudios y que también representan una diferencia con PISA, es que se conceptualizan y miden variables a nivel de aula y de profesor, pudiéndose identificar dimensiones explicativas asociadas a la mediación pedagógica y a la dinámica de la clase. Otra ventaja de estos estudios es que permiten medir factores idiosincráticos del contexto del país que pueden no ser relevantes a nivel internacional y no aparecer en los cuestionarios de PISA.

Además, estas pruebas se realizan en las cinco asignaturas principales (Matemática, Español, Estudios Sociales, Ciencias e Inglés) a diferencia de las pruebas internacionales que tienden a concentrarse solo en Lenguaje, Matemática y Ciencias.

Finalmente, los constructos objeto de medición en las pruebas diagnósticas del MEP son conocimientos y destrezas específicos que se esperan como resultados directos del proceso de aprendizaje y que se derivan del currículo de cada materia, por tanto son más apropiadas como

termómetro del logro de aprendizajes académicos formales establecidos en los programas de estudios, a diferencia de las pruebas PISA que no se ajustan a ningún currículo en particular.

Entre algunas de las variables que más consistentemente generan evidencia de asociación directa con los rendimientos en estas pruebas se pueden mencionar las siguientes: el ser alumno de un colegio privado, poseer un mayor nivel socioeconómico, nivel de atención y participación en el aula, motivación de la familia y satisfacción del director con el clima escolar. Los estudiantes varones presentan diferencias estadísticamente significativas a su favor al compararlos con las mujeres en Matemática, Estudios Sociales e Inglés. El grado académico del docente se asocia significativamente a mayores rendimientos en Matemática e Inglés. (MEP, 2012).

El índice socioeconómico se relacionó significativamente con el rendimiento académico del estudiantado de noveno año en Matemática. Cabe destacar que este resultado coincide con estudios internacionales realizados por la UNESCO en América Latina (Oviedo Y., 2012). Asimismo la comparación de medias por zona indicó con un 95% de confianza que sí existen diferencias entre las medias de los puntajes de los examinados de colegios ubicados en la zona urbana y aquellos que se localizan en la zona rural, siendo la diferencia a favor de los urbanos. Este estudio también encontró que la variable sector (público-privado) está relacionada significativamente con el rendimiento académico en Matemática, presentando una magnitud bastante alta en el coeficiente beta, indicador de la importancia práctica del resultado (Oviedo, 2012).

3.2.4. Estudios que utilizan datos de PISA para Costa Rica

La prueba PISA se aplica desde el año 2000, y aunque Costa Rica sólo ha participado en cuatro de las siete evaluaciones realizadas hasta la fecha, existen algunos trabajos que estudian los factores asociados al rendimiento educativo utilizando las bases de datos de PISA y que involucran a nuestro país.

Los dos primeros estudios publicados que se enfocan en la participación de Costa Rica en la prueba PISA 2009 fueron los de Fernández y Del Valle (2013) y Montero y Otros (2013). En el primero de estos, los autores intentan responder a la pregunta de si la desigualdad en los resultados de los puntajes en la prueba PISA se deben en gran medida a las diferencias entre los estudiantes que asisten a colegios públicos o a colegios privados.

Para intentar responder a esta pregunta, los autores plantean un modelo de regresión lineal sin corrección multinivel, con el objetivo de poder aplicar una técnica de descomposición de la varianza que permita identificar si la proporción de varianza explicada por la variable del sector del colegio (controlando por otros factores que describe la teoría como importantes) es significativa en la explicación de las diferencias en los resultados educativos.

Sus resultados muestran que, en general, los factores escolares (incluyendo el sector del colegio) no tienen un gran peso en la variabilidad de los resultados, y más bien son los factores familiares, pero particularmente la condición de repitencia de los estudiantes costarricenses, los que más explican la variabilidad en la varianza de las puntuaciones.

En un estudio similar, Giménez y otros (2014) plantean la pregunta de ¿por qué los estudiantes de colegios públicos y privados de Costa Rica obtienen distintos resultados académicos? Para dar respuesta, llevan a cabo un análisis econométrico que, mediante la utilización de funciones de producción educativa y estimaciones por mínimos cuadrados ordinarios con errores estándar corregidos utilizando clústeres por centros educativos, permite cuantificar las diferencias en resultados en pruebas de Matemáticas, Comprensión Lectora y Ciencias entre centros públicos y privados. Aunque las diferencias halladas son importantes, se reducen en más de la mitad cuando se controla por las características de los alumnos (sexo, haber cursado preescolar, confianza en sus capacidades, ser repetidor) y sus hogares (nivel de estudio de los padres, riqueza y posesiones culturales).

Además, mediante el método de descomposición de Oaxaca-Blinder (Oaxaca, 1973), en ese estudio se profundiza en cuáles son las posibles razones de estas diferencias en resultados. Concluyen que la mayor parte de las diferencias en resultados entre centros públicos y privados es explicada por la diferencia en dotaciones. En concreto, el principal factor de la divergencia en rendimiento son las características del hogar, seguida de las diferencias en recursos de los centros educativos, las características del estudiante y el ambiente de trabajo de los centros educativos.

Montero y otros (2013) utilizan modelos de regresión con corrección multinivel con la base de datos de PISA 2009. logran demostrar que las variables relacionadas con dimensiones de lectura, tienen, conjuntamente, un peso altísimo en la predicción de las notas, tanto en Competencia Lectora como en Alfabetización Matemática. De acuerdo con los resultados de este estudio, aproximadamente un 43% de la varianza de los puntajes se explica por factores asociados al estudiante y un 12% por factores asociados a la institución.

De hecho, el estudio va más allá y elabora un perfil de un típico estudiante de alto rendimiento, que incluye lo siguiente: cursa un grado mayor al que debería estar de acuerdo con su edad, mejor actitud hacia la lectura, usa mejores estrategias para comprender y resumir textos escolares, no es repitente, valora su colegio, lee por gusto o por iniciativa propia, su hogar tiene un alto índice de posesión de bienes, entre otros. El único factor relacionado con la institución que aparece en este perfil es asistir a un centro educativo ubicado en un distrito con altos valores del índice de desarrollo social.

Montero y otros (2014) vuelven a analizar los resultados de la prueba PISA con un modelo de regresión con corrección multinivel y de la misma manera fueron identificadas variables que son importantes en la predicción de los puntajes de las tres pruebas de PISA y que, a su vez, se considera que pueden ser objeto de intervenciones pedagógicas o de políticas educativas, con el objetivo de lograr mejores rendimientos en dichas pruebas. Encuentran que el número de libros en el hogar es una de las variables predictoras importantes para la Alfabetización Matemática, puesto que se puede considerar un indicador *proxy* de dimensiones de lectura, las cuales fueron, conjuntamente, el factor predictivo más importante para el puntaje en las pruebas PISA del año 2009.

Además, otros factores con gran importancia predictiva son el número de horas semanales que se dedica a la realización de tareas, un alto valor en la escala de familiaridad con conceptos

matemáticos, un alto valor en la escala de autoeficacia en matemáticas, un alto valor en la escala de actitud hacia lo aprendido en el colegio, el haber asistido a educación preescolar y un alto valor en la cantidad de docentes de matemáticas con bachillerato o licenciatura en pedagogía.

Por su parte, en un primer intento por plantear un modelo de ecuaciones estructurales con datos de la prueba PISA 2009 para Costa Rica, Montero (2012) realiza un ejercicio ilustrativo para ejemplificar el potencial de este enfoque de análisis de datos en investigación educativa, utilizando algunas de las variables del estudiante y del centro educativo que resultan más relevantes para su predicción. Con ellas se desarrollaron diversos ensayos para la estimación de modelos estructurales que tuvieran sentido teórico, que pudieran estimarse de manera técnicamente correcta y que además arrojaran niveles relativamente aceptables de ajuste estadístico.

Dentro de los resultados más destacables (Montero, 2012) están los siguientes: el constructo extracción social tiene una relación causal importante con el capital cultural y con el rendimiento en la prueba, mientras que encuentra una relación inversa a la esperada entre extracción social y actitudes y hábitos hacia la lectura.

Según la autora, la magnitud de la relación entre antecedentes educativos y la nota en la prueba PISA de competencia lectora parece razonable, mientras que la relación entre el constructo estrategias de lectura (comprender y resumir un texto) y actitudes y hábitos hacia la Lectura también resulta significativa, al igual que la relación estimada entre estas estrategias y la nota en la prueba.

Montero (2012) encuentra evidencia una fuerte relación entre capital cultural y las actitudes y hábitos hacia la lectura. El capital cultural tiene así un efecto indirecto de importancia sobre el rendimiento en la prueba de PISA de competencia lectora, a través de actitudes y hábitos de lectura, pero encuentra que el efecto directo no parece tener importancia práctica.

Fernández (2013) analiza a los estudiantes costarricenses identificados como “resilientes” en la evaluación PISA 2009; y los compara con sus pares que, a pesar de tener un nivel socioeconómico similar, obtuvieron malos resultados en esta prueba. La resiliencia educativa refiere a los estudiantes que, a pesar de vivir bajo condiciones socioeconómicas que les son adversas, lograron obtener excelentes resultados en la evaluación internacional PISA 2009 y así romper el círculo vicioso que hay entre un pobre entorno socioeconómico y un bajo rendimiento educativo. Mediante el uso de una regresión logística, el modelo fue estimado utilizando tres variables de control: el género, el grado que cursa el estudiante y el promedio del índice estatus económico social y cultural a nivel de centro educativo (efecto de pares). En Montero (2012) se identifican cuatro factores principales que podrían estar relacionados con el mejor desempeño en lectura de los estudiantes resilientes: 1) la habilidad para resumir textos, 2) la habilidad para entender y recordar, 3) el gusto por la lectura y 4) la actitud hacia el colegio.

Los dos primeros factores se relacionan con el tipo de estrategias que utilizan los estudiantes para comprender un texto y para transmitir la información aprendida mientras que los dos últimos

factores se relacionan más el con valor que los estudiantes le confieren a la lectura y a los deberes académicos.

Finalmente, Fernández (2015) evalúa la hipótesis causal de que el estatus socioeconómico de los estudiantes y la actitud de los estudiantes hacia la matemática son factores que determinan en gran medida los resultados académicos de los estudiantes costarricenses, medido a través del resultado en la prueba de alfabetización matemática de PISA 2012. Para esto, define un modelo de medición de los constructos latentes y estima un modelo de ecuaciones estructurales.

El estatus socioeconómico de los estudiantes es medido mediante tres variables indicadoras: el índice posesiones del hogar, el índice de estatus ocupacional más alto de los padres, y el índice del nivel educativo más alto de los padres.

La actitud de los estudiantes hacia las matemáticas es una variable latente que intenta capturar la actitud propia de los estudiantes hacia el aprendizaje y la utilización de las matemáticas. Este factor es medido mediante la utilización de tres variables indicadoras: la confianza en la realización de problemas matemáticos, la ansiedad a la hora de enfrentarse a problemas que involucren la utilización de matemáticas y el auto concepto, que mide qué tan bueno o malo se considera el estudiante a sí mismo para las matemáticas.

Fernández (2015) concluye que el modelo estructural brinda bastante evidencia a favor de la hipótesis de causalidad de que el estatus socioeconómico y la actitud personal del estudiante hacia las matemáticas son determinantes del rendimiento educativo de los estudiantes costarricenses en esta materia, medido a través de los resultados de la prueba de alfabetización matemática de PISA 2012, siendo este un buen punto de partida para formular un modelo más extenso que contemple de mejor manera la complejidad de los factores sociales, institucionales y de contexto que inciden en el rendimiento académico de los jóvenes.

3.2.5. Síntesis del capítulo

Como fue explicado, en la literatura internacional no existe un claro consenso sobre cuáles factores son los que inequívocamente se asocian con un mejor rendimiento educativo de los estudiantes. Una posible explicación es que lo que funciona para un país no necesariamente debe funcionar en otro, por las particularidades del contexto educativo, político, económico y social de cada uno.

Dadas las particularidades del efecto de cada intervención dependiendo del contexto en el que fue aplicado, resulta necesario el tomar en cuenta con mayor énfasis la revisión de literatura para el caso de Costa Rica, ya que estas investigaciones sí incorporan dentro de sus marcos de referencia el contexto en el que se desenvuelve nuestro sistema educativo. Sin embargo, dado que la mayoría de las investigaciones analizadas en este capítulo basan sus resultados en modelos de regresión que estiman efectos directos, y no explican posibles trayectorias de efectos indirectos entre factores (que son vitales en los modelos de ecuaciones estructurales), una tarea adicional es la de plantear estas relaciones.

Por lo tanto, con base en esta información y en el propio marco de referencia sobre alfabetización matemática de PISA, fue planteado un modelo teórico de relaciones entre factores que inciden en el rendimiento cognitivo. Este modelo teórico fue puesto a discusión y validación por parte de expertos en el tema mediante sesiones de trabajo realizadas en el año 2016. La descripción completa del modelo se encuentra en la sección 4.2.5.

4. Abordaje metodológico

El objetivo de esta investigación es generar un marco comparativo del desempeño de modelos de ecuaciones estructurales para datos multinivel con estimación Bayesiana, con el propósito de tomar decisiones en la modelación del caso aplicado. Para ello, fueron generados datos simulados que permitieron ensayar diversidad de condiciones de análisis y así observar su efecto en las estimaciones de estos modelos. Los resultados obtenidos sirvieron como base para identificar la manera adecuada de analizar la relación entre las variables asociadas el rendimiento educativo de los estudiantes y su nivel de alfabetización matemática, en un estudio con las características particulares de PISA.

Es decir, la investigación se dividió en dos partes:

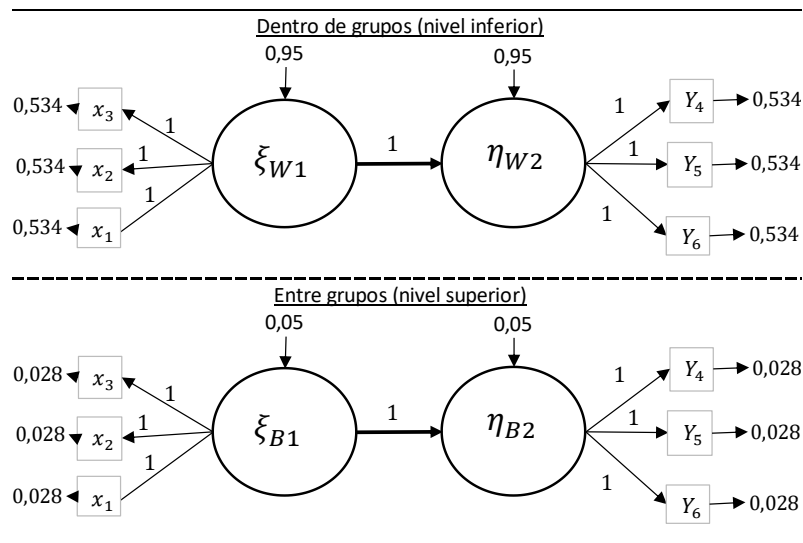
- Evaluación simulada de la robustez de la estimación ML-BSEM.
- Análisis de datos de la evaluación PISA de acuerdo con los resultados de la primera parte.

4.1. Evaluación simulada de la robustez del modelo ML-BSEM

4.1.1. Descripción del modelo

El modelo SEM multinivel que se utilizó en las simulaciones reportadas en esta investigación consiste en una covariable latente y una variable respuesta latente, cada una de las cuales es medida por tres variables indicadoras. La escogencia de esta estructura busca mantener comparabilidad con la línea de investigación en simulaciones realizadas en estudios precedentes (Lüdtke, Marsh, Robitzsch, & Trautwein, 2011), (Rabe-Hesketh, Skrondal, & Pickles, 2004), (Depaoli & Clifton, 2015)). La siguiente figura muestra el modelo para una de las condiciones que serán explicadas seguidamente:

Figura 4 Modelo de generación de datos con coeficiente de correlación intra-clase de 0.05



En el enfoque en dos etapas (ver sección 2.3) la estructura de la covariación es dividida en dos componentes: el componente *dentro de grupos*, y el componente *entre grupos*. De esta forma, son estimados modelos separados para cada uno de estos componentes.

El componente “*dentro de grupo*” (que será denotado con el subíndice W) representa la variación al nivel de la unidad más baja, mientras que el componente “*entre grupos*” (denotado por el subíndice B) representa la variación entre los grupos del nivel superior.

Para la observación i y el grupo j , el vector de respuesta y_{ij} se puede descomponer como:

$$y_{ij} = \mu + y_{Wij} + y_{Bj}$$

En donde los componentes “*dentro de grupo*” (y_{Wij}) y “*entre grupos*” (y_{Bj}) son ortogonales e independientes, y μ representa las grandes medias. El vector y_{ij} está normalmente distribuido con medias entre grupos μ_j (interceptos aleatorios) y matriz de covarianzas Σ_W .

Los efectos aleatorios μ_j se distribuyen normalmente con el valor esperado μ y la matriz de covarianzas Σ_B .

Para la estimación del modelo referido al componente “*dentro de grupo*”, el modelo de medición es:

$$y_{ij} = \mu_j + \Lambda_W \eta_{Wij} + \varepsilon_{Wij}$$

Donde Λ_W es una matriz de cargas factoriales de tamaño 6×2 , que en el modelo tradicional incluiría cargas factoriales de cero para las cargas cruzadas, pero que fue parte de las variaciones planteadas en el diseño. Esta matriz es post multiplicada por un vector de tamaño 2×1 de variables latentes del primer nivel η_{Wij} que contiene la variable respuesta latente y la covariable latente.

El vector η_{Wij} se distribuye mediante una normal multivariada con una esperanza de cero y una matriz de covarianzas de tamaño 2×2 Ψ_W . Dado que las variables latentes no están correlacionadas (o al menos eso se supondrá), la matriz Ψ_W contiene las varianzas de los factores a lo largo de la diagonal, y ceros en el resto de entradas (una matriz diagonal).

El vector de tamaño 6×1 que contiene los errores, ε_{Wij} , se distribuye con una normal multivariada con una esperanza de cero y una matriz de covarianzas diagonal Θ_W , con términos de error a lo largo de la diagonal.

El modelo estructural para el componente “*dentro de grupos*” tiene la siguiente forma:

$$\eta_{Wij} = B_W \eta_{Wij} + \zeta_{Wij}$$

Donde B_W es una matriz de tamaño 2×2 de coeficientes estructurales definida como:

$$\begin{bmatrix} 0 & b_W \\ 0 & 0 \end{bmatrix}$$

Donde b_W representa la regresión del factor latente respuesta sobre la covariable latente (el parámetro estructural del modelo SEM). La matriz B_W es multiplicada por el vector de variables

latentes del primer nivel para obtener la estimación del efecto de la covariable. Finalmente, ζ_{wij} es un vector de tamaño 2×1 que contiene los errores, que se distribuyen normal multivariado con una esperanza de cero y una matriz de covarianzas diagonal Ω_W que incluye términos de error en la diagonal.

Los modelos de medición y estructural para el componente “entre grupos” (el nivel superior jerárquico) se representan, respectivamente, de la siguiente forma:

$$\begin{aligned}\mu_j &= \mu + \Lambda_B \eta_{Bj} + \varepsilon_{Bj} \\ \eta_B &= B_B \eta_{Bj} + \zeta_{Bj}\end{aligned}$$

4.1.2. Especificación de las distribuciones a priori

El grado con el que los parámetros de un análisis Bayesiano son recuperados con precisión depende en gran medida de la calidad y la cantidad de información modelada en la distribución a priori del mismo.

Existen tres grandes categorías de distribuciones a priori que son discutidas típicamente en términos de su nivel de información: no informativas (o difusas), débilmente informativas e informativas. Las distribuciones a priori difusas contribuyen con casi nula información al modelo; las distribuciones a priori débilmente informativas contienen más información que las difusas, pero utilizan menos información de la que está disponible y aún mantienen cierto grado de incertidumbre con respecto al parámetro. Finalmente, las distribuciones a priori informativas incluyen una gran cantidad de información y certidumbre acerca del valor del parámetro del modelo¹⁵. Un análisis Bayesiano exitoso requiere de una selección cuidadosa de las distribuciones a priori, incluso si son no informativas.

Las distribuciones a priori se cuantifican con la especificación de hiperparámetros que generalmente controlan la ubicación y la escala de la distribución. Por ejemplo, una distribución a priori para una variable normalmente distribuida tiene un hiperparámetro para la media y un hiperparámetro para la varianza que corresponden a la ubicación y a la escala de la distribución, respectivamente.

Es probable que el hiperparámetro de la varianza para una distribución a priori no informativa sea mucho mayor que el de una distribución a priori informativa para reflejar un mayor grado de incertidumbre sobre el rango de valores que puede asumir el parámetro.

Con muestras de gran tamaño, las distribuciones a priori no informativas tienen poco impacto en la inferencia de la distribución a posteriori ya que la verosimilitud toma más relevancia. Esto implica que las estimaciones de modelos para muestras pequeñas son más sensitivas a la especificación de la distribución a priori.

Para el presente modelo, los parámetros de interés incluyen los parámetros de regresión y de varianza en cada nivel del modelo. Seguidamente se explicarán las distribuciones especificadas en cada parámetro. En la notación se muestran como multivariadas, pero es importante aclarar que el

¹⁵ Las distribuciones a priori débilmente informativas pueden ser preferidas a las informativas cuando no es claro cómo construir un modelo completamente informativo y por lo tanto es apropiado permitir a la verosimilitud contribuir de mayor forma que la distribución a priori (Depaoli & Clifton, 2015).

software *Mplus* implementa distribuciones a priori univariadas en los elementos individuales de un vector (Muthén & Muthén, (1998-2015)).

La distribución a priori para el vector de parámetros de regresión es la distribución normal multivariada. Para los parámetros de regresión del componente “*dentro de grupos*” $\delta_W = (\Lambda_W, \mathbf{b}_W)$, la distribución a priori se denota como:

$$\delta_W \sim MNV(v_W, \sigma_W^2)$$

donde v_W es un vector que contiene los hiperparámetros de la media y σ_W^2 es un vector que contiene los hiperparámetros para la varianza.

De la misma forma, las distribuciones a priori para los parámetros de regresión del nivel de grupos $\delta_B = (\Lambda_B, \mathbf{b}_B)$ se definen como:

$$\delta_B \sim MNV(v_B, \sigma_B^2)$$

donde v_B y σ_B^2 representan los vectores que contienen los hiperparámetros de la media y la varianza, respectivamente.

La distribución a priori especificada para la varianza en la Gamma-Inversa (Γ^{-1}), que incluye parámetros de forma (α) y de escala (β) que controlan la densidad.

También son establecidas distribuciones Gamma-Inversa (Γ^{-1}) para los elementos de la diagonal de las matrices de covarianzas asociadas con cada varianza en cada nivel del modelo. En el nivel “*dentro de grupo*”, las distribuciones a priori para cada uno de los parámetros de varianza son:

$$\Theta_W \sim \Gamma^{-1}(\alpha_{W_1}, \beta_{W_1}) \quad \text{y} \quad \Psi_W \sim \Gamma^{-1}(\alpha_{W_2}, \beta_{W_2})$$

donde α_W es el hiperparámetro de forma y β_W es el hiperparámetro de escala.

De forma análoga, la distribución a priori para cada parámetro de varianza del nivel “*entre grupos*” se denota como:

$$\Theta_B \sim \Gamma^{-1}(\alpha_{B_1}, \beta_{B_1}) \quad \text{y} \quad \Psi_B \sim \Gamma^{-1}(\alpha_{B_2}, \beta_{B_2})$$

4.1.3. Diseño de las simulaciones

Los datos fueron generados utilizando la versión Mplus 8.3¹⁶ (Muthén & Muthén, (1998-2015)) con 1.000 repeticiones por celda¹⁷. El análisis Bayesiano fue implementado utilizando una cadena MCMC de 50.000 iteraciones cada una descartando las primeras 25.000, y utilizando las restantes 25.000 para estimar la distribución a posteriori, aunque la convergencia ocurrió mucho antes en la mayoría de estimaciones.

La convergencia de la cadena fue monitoreada con el diagnóstico de convergencia de Brooks y Gelman (1998), que es el implementado en *Mplus*, con el criterio de convergencia por defecto de 0.05.

El modelo para la generación de datos (parámetros simulados) consistió en una variable latente explicatoria y una variable latente respuesta tanto en el nivel de individuos como en el de grupos, donde cada variable latente es medida por tres variables indicadoras continuas.¹⁸ Para los propósitos de identificación del modelo, la primera variable indicadora de cada variable latente fue fijada al valor de 1.0 en todas las condiciones.

Para el modelo multinivel de dos niveles en la estructura jerárquica, las condiciones que fueron modificadas en este estudio son el número de grupos (2 variantes), el tamaño de los grupos (2 variantes), el coeficiente de correlación intra-clase (3 variantes) y las especificaciones de las distribuciones a priori de las regresiones y las varianzas del modelo (6 variantes).

Seguidamente se explican estas variantes:

Número de grupos (2 variantes): Las investigaciones Monte Carlo han reportado diferentes hallazgos con respecto al número de grupos que son necesarios para obtener estimaciones insesgadas en ML-SEM (Li & Beretvas, 2013; Ludtke y otros, 2011; Preacher y otros, 2010; Depaoli & Clifton, 2015). En general, la estimación Bayesiana produce aceptables resultados incluso para un número de grupos pequeño (Depaoli & Clifton, 2015). Debido a lo anterior, no fue especificado un amplio rango de grupos a evaluar, sino únicamente dos, que podrían clasificarse en pequeño número de grupos y gran número de grupos, para capturar la dispersión al variar esta categoría. Los grupos son de *tamaño 30 y 100* (en la evaluación PISA 2012, el número de grupos fue de 193).

Tamaño del grupo (2 variantes): Para cada grupo, *tamaños de muestra de 10 y 30* observaciones fueron generados para reflejar los tamaños de grupos típicamente observados en estudios

¹⁶ La versión 8.3 fue puesta en producción en abril de 2019. En lo que concierne a esta investigación, esta versión añade ciertas funcionalidades con respecto a la versión 8.2, que colaboran en los procesos de estimación de modelos Bayesianos multinivel, especialmente en los enfoques de computación paralela y la reorganización de algoritmos de estimación Bayesiana, que reducen los tiempos de uso computacional entre 20% y 50% (Asparouhov & Muthén, 2019).

¹⁷ La celda es el script o código utilizado para la especificación y estimación del modelo, en el lenguaje de MPlus.

¹⁸ Inicialmente se planteó la posibilidad de realizar también simulaciones tomando en cuenta constructos creados a partir de variables categóricas, específicamente variables indicadoras dicotómicas. Sin embargo, debido al tiempo computacional requerido para las simulaciones fue descartada esta aplicación.

educativos (en la evaluación PISA 2012 el tamaño de grupo estándar es de 30 estudiantes). De la misma forma, la idea es evaluar el efecto de contar con pocas observaciones dentro de cada grupo, o con gran cantidad de observaciones.

Al combinar las variantes anteriores, se obtienen cuatro tamaños de muestra total, a saber: 300 observaciones, 900 observaciones, 1000 observaciones y 3000 observaciones (el tamaño de muestra total de la evaluación PISA 2012 es de 4602 observaciones para Costa Rica).

Coefficiente de correlación intra-clase (3 variantes): La descomposición de la varianza en los componentes “dentro de grupos” y “entre grupos” es un factor que interactúa junto con el tamaño de la muestra y que puede afectar tanto los resultados de las estimaciones como la convergencia misma, especialmente en las estimaciones frecuentistas. El coeficiente de correlación intra-clase (ICC) es un índice de la proporción de la variabilidad que es explicada por el nivel de grupos:

$$\rho = \frac{Var(B)}{Var(B) + Var(W)}$$

La correlación intraclase mide la homogeneidad interna de los grupos, en otras palabras, la similitud de las unidades de primer nivel y las diferencias entre las unidades de niveles superiores. Un valor de cero indica que los datos son independientes, pero entre mayor sea mayor importancia adquiere el tomar en cuenta la estructura jerárquica al momento de realizar estimaciones de regresión.

En los estudios relacionados al sector educación, es común encontrar ICC sumamente pequeños (Fernández, 2017), por lo que fueron simulados *tres escenarios del valor del ICC: 0.02, 0.05 y 0.40*.

En los modelos multinivel, el ICC para cada variable observada (ρ_k) puede ser expresado como una función de las cargas factoriales, varianzas, y varianzas residuales en cada nivel del modelo utilizando la siguiente fórmula (Muthén B. , 1991):

$$\rho_k = \frac{(\lambda_{Bk}^2 \times \Psi_B + \Theta_B)}{[(\lambda_{Bk}^2 \times \Psi_B + \Theta_B) + (\lambda_{Wk}^2 \times \Psi_W + \Theta_W)]}$$

donde λ_k se refiere a la carga factorial del ítem k .

Los ICC que fueron utilizados en esta investigación fueron creados utilizando la ecuación anterior, al cambiar las varianzas y las varianzas residuales de cada nivel del modelo manteniendo las cargas factoriales en un valor constante de 1.0.

Para obtener un ICC de 0.02 en el modelo con datos continuos, se establece la varianza factorial del nivel bajo (“dentro de grupos”) en un valor de 0.98, las varianzas residuales del nivel bajo en 1.74, la varianza factorial del nivel superior en un valor de 0.02, y las varianzas residuales del nivel superior en un valor de 0.035.

Para obtener un ICC de 0.05 en el modelo con datos continuos, se establece la varianza factorial del nivel bajo (“dentro de grupos”) en un valor de 0.95, las varianzas residuales del nivel bajo en 0.534, la varianza factorial del nivel superior en un valor de 0.05, y las varianzas residuales del nivel superior en un valor de 0.028.

Para obtener un ICC de 0.40 para el modelo con datos continuos, se establece la varianza factorial del nivel bajo (“dentro de grupos”) en un valor de 0.60, las varianzas residuales del nivel bajo en 0.3375, la varianza factorial del nivel superior en un valor de 0.40, y las varianzas residuales del nivel superior en un valor de 0.225.

Como referencia, la figura 1 anterior muestra los valores de los parámetros para la celda con el ICC de 0.05.

Al combinar los cuatro tamaños de muestra con los tres contextos en los que se probará el efecto del Coeficiente de Correlación Intra-clase, se producen un total de doce combinaciones, que son llamadas en esta investigación como los doce escenarios de contexto, ya que si bien permiten realizar comparaciones del desempeño de las estimaciones bajo estos escenarios, por si solas no son útiles para responder a la pregunta de investigación, que está relacionada con la especificación de las distribuciones a priori de algunos parámetros, y que será comentado a continuación.

Especificaciones de las distribuciones a priori (6 variantes): La estimación Bayesiana fue realizada utilizando cadenas MCMC con el algoritmo de muestreo de Gibbs (ver secciones 2.4.3 y 2.4.4). Tres tipos de distribuciones a priori fueron especificadas para todo el modelo: no informativa, informativa y débilmente informativa, siendo esta última definida como una distribución con un nivel de variación mayor al de las distribuciones informativas pero inferior al de las no informativas.

Los parámetros de regresión fueron especificados para tener la misma distribución a priori entre los dos componentes del modelo (“dentro de grupo” y “entre grupo”).

Los parámetros de varianza para el componente “dentro de grupo” del modelo fueron siempre especificadas con los elementos en $\tau_W \sim \Gamma^{-1}(-1, 0)$, que es el valor por defecto en *Mplus* y corresponde a la distribución uniforme $U[0, \infty)$ (Asparouhov & Muthén, 2010).

Las distribuciones a priori difusas (no informativas) fueron definidas utilizando las especificaciones por defecto que incluye el programa *Mplus*.

Para la estimación Bayesiana con distribuciones a priori informativas y débilmente informativas, el hiperparámetro de media de las regresiones fue especificado al verdadero valor poblacional (establecido en la generación de datos) de 1.0 para darle a la distribución a priori de la regresión la mayor masa en la vecindad general de los valores generados (Depaoli & Clifton, 2015).

En el caso de las distribuciones a priori débilmente informativas, fueron especificados dos niveles de información para las regresiones a través del hiperparámetro de varianza de tal forma que $\delta_W \sim N(1, 1)$ y $\delta_W \sim N(1, 0.5)$ y de la misma forma, $\delta_B \sim N(1, 1)$ y $\delta_B \sim N(1, 0.5)$. Los componentes de varianza en el nivel jerárquico superior fueron modificados para tener distribuciones a priori $\Gamma^{-1}(-1, 0)$ o $\Gamma^{-1}(0.01, 0.01)$. Para estudiar la interacción de las diferentes formas de las distribuciones a priori débilmente informativas, las especificaciones de las regresiones y de las varianzas fueron cruzadas, dando como resultado cuatro niveles de definiciones de distribuciones a priori débilmente informativas.

Con respecto a las distribuciones a priori informativas, el hiperparámetro de varianza de las distribuciones normales fue especificado para ser más pequeño que las distribuciones a priori débilmente informativas recién comentadas, de forma que $\delta_W \sim N(1, 0.1)$ y $\delta_W \sim \Gamma^{-1}(0.001, 0.001)$. Lo mismo aplica para el nivel jerárquico superior, $\delta_B \sim N(1, 0.1)$ y $\delta_B \sim \Gamma^{-1}(0.001, 0.001)$.

En síntesis, para la estimación de los parámetros de regresión y de varianza (excluyendo, de momento, las cargas factoriales cruzadas) fueron testeadas seis combinaciones de distribuciones a priori: una no informativa, 4 débilmente informativas, y una informativa.

Cargas factoriales cruzadas (distribuciones a priori) (3 variantes): La pregunta de investigación refiere al estudio de las posibles ventajas del enfoque BSEM en el contexto de los modelos de ecuaciones estructurales multinivel (ML-BSEM). Para dar respuesta a la pregunta, fue introducida una variante adicional en el diseño de la simulación: las matrices Λ_W y Λ_B , que contienen las cargas factoriales y los parámetros de regresión - y que fueron modeladas con distribuciones a priori distribuidas normalmente con tres variaciones en el hiperparámetro de la varianza- fueron modificadas para permitir simultáneamente la libre estimación de los parámetros de las cargas factoriales cruzadas, que originalmente, en el diseño mostrado en la sección 4.1.3, serían fijados a un valor de 0 ($\sim N(0, 0)$).

En este caso, para cada uno de los 72 modelos de base (12 escenarios de contexto combinados con 6 especificaciones de las distribuciones a priori en los parámetros de regresión y de varianza) las cargas factoriales cruzadas fueron estimadas con una distribución a priori normal que tendrán una media de cero y una varianza aproximadamente igual a cero (el enfoque BSEM).

Se recurrió al uso de tres variantes para medir el efecto de las cargas factoriales cruzadas: $\sim N(0, infinito)$, $\sim N(0, 1)$ y $\sim N(0, 0.1)$, para evaluar el efecto de una distribución a priori no informativa, débilmente informativa e informativa, siguiendo el enfoque BSEM (Muthén & Asparouhov, 2012 (1)). De esta forma, fueron especificados 216 modelos (celdas) con variable indicadora continua.

4.1.4. Evaluación del ajuste del modelo

Fueron realizadas simulaciones sobre la combinación de distintas condiciones que pueden afectar tanto la convergencia del modelo como la correcta estimación o recuperación de los parámetros simulados. El diseño es totalmente cruzado, lo que da como resultado un total de doce escenarios de condiciones contextuales, que al añadirseles las seis distintas combinaciones para establecer las distribuciones a priori de los parámetros principales, da un total de 72 celdas para las simulaciones del modelo SEM multinivel, que al añadir las tres especificaciones para las cargas factoriales cruzadas, alcanzan un total de 216 celdas. A cada una de las 216 celdas le fueron simuladas 1000 bases de datos y fueron estimadas las 1000 repeticiones correspondientes, lo que implica que el modelo fue estimado 216 mil veces.

Los objetivos específicos de esta investigación se basan en tres evaluaciones a los modelos: a) la estabilidad de las estimaciones de los parámetros, b) la cantidad de sesgo introducida en las estimaciones de los parámetros, y c) el ajuste de los modelos mediante el cómputo de la

probabilidad asociada a la prueba cociente de verosimilitudes del modelo nulo respecto del modelo irrestricto.

Para concretar la evaluación resultó necesario definir algún indicador mediante el cual pudieron ser examinados los cambios en los modelos (en términos de sesgos, estabilidad de parámetros y ajuste). Para la estabilidad de las estimaciones de los parámetros, fueron examinadas ciertas medidas de posición de la distribución de las estimaciones producidas en las simulaciones y la raíz cuadrada del error cuadrático medio. Para la cantidad de sesgo en las estimaciones, fue calculada la diferencia entre el valor dado del parámetro y el valor estimado, de forma que se puedan apreciar las desviaciones, o no, del valor de la pendiente desde el cual se simulan los datos. Finalmente, el cambio en el ajuste de los modelos se evaluará mediante el cómputo de la probabilidad asociada a la prueba cociente de verosimilitudes del modelo nulo respecto del modelo irrestricto, haciendo uso del *P-value predictivo a posteriori*.

El *P-value predictivo a posteriori* del ajuste del modelo puede obtenerse con un estadístico de ajuste, f , basado en el test usual del cociente de verosimilitudes de un modelo H_0 con respecto a un modelo irrestricto H_1 . En la implementación de Mplus, el *P-value predictivo a posteriori* es computado utilizando cada décima iteración entre las iteraciones utilizadas para describir la distribución a posteriori de los parámetros (Asparouhov & Muthén, 2010).

Se construye un intervalo al 95% de confianza para la diferencia en el estadístico f entre los datos reales y replicados. Un límite inferior positivo va en línea de un bajo *P-value predictivo a posteriori* e indica un pobre ajuste. Un modelo con ajuste excelente se espera que tenga un *P-value predictivo a posteriori* alrededor de 0.5 y una diferencia en el estadístico f de cero que se ubique cerca del punto medio del intervalo de confianza (Muthén & Asparouhov, 2012 (1)).

Es importante destacar que el *P-value predictivo a posteriori* no se comporta como un p-value para una prueba chi-cuadrado de ajuste de modelos. La probabilidad de error Tipo I no es 0.05 para un modelo correcto, y aún no hay teoría para qué tan bajo debe ser el valor del *P-value predictivo a posteriori* antes de que el modelo pase a mostrar un mal ajuste a un cierto nivel. En ese sentido, el *P-value predictivo a posteriori* es más similar a un índice de ajuste de un SEM (como el RMSEA) que a una prueba chi-cuadrado (Muthén & Asparouhov, 2012 (1))¹⁹.

De esta forma, mediante el uso del *P-value predictivo a posteriori* fue realizada la comparación del ajuste de los modelos: para cada uno de los 48 escenarios de contexto propuestos, las 12 celdas correspondientes fueron comparadas (6 que no incluyen distribuciones a priori en las cargas factoriales cruzadas y 6 que sí lo incluyen) para evaluar las bondades de utilizar el enfoque ML-BSEM²⁰.

¹⁹ Si bien el uso del *P-value predictivo a posteriori* ha sido debatido recientemente en cuanto a su robustez para evaluar y comprar el ajuste de modelos con estimación Bayesiana (Hojtink & van de Schoot, 2017), el hecho de que aún la evidencia no sea contundente para descartarlo lo convierte en el indicador que será utilizado en esta investigación para comparar el ajuste de los modelos.

²⁰ En noviembre de 2019 Asparouhov y Muthén (2019) publicaron un nuevo documento que muestra actualizaciones al PPP-value en presencia de datos faltantes, así como una adaptación Bayesiana de los índices de ajuste de aproximación tradicionalmente utilizados en modelos SEM, como el RMSEA, CFI y el TLI. Estos fueron agregados a la versión 8.4 de Mplus, que fue liberada en noviembre 2019, por lo que no pudo ser aplicado en esta investigación (además de que aún no se encuentran disponibles para modelos multinivel).

4.2. Análisis de los datos de la evaluación PISA

4.2.1. El Programa Internacional de Evaluación de Estudiantes (PISA)

El Programa Internacional de Evaluación de Estudiantes (*PISA- Programme for International Student Assessment*) es un esfuerzo conjunto entre los países miembros de la OCDE que busca medir la capacidad que tienen los estudiantes en edad de 15 años y que se acercan a concluir el ciclo de escolaridad obligatoria, para afrontar los retos de la sociedad actual.

La evaluación tiene una visión a futuro: en lugar de centrarse en evaluar conocimientos específicos de acuerdo con los currículos o planes de estudio de cada centro educativo, se evalúa la capacidad que tienen los estudiantes para utilizar su conocimiento y habilidades para afrontar retos de la vida real.

Las encuestas y evaluaciones de PISA se realizan cada tres años y aunque todas miden habilidades y conocimientos en lectura, ciencias y matemáticas, cada evaluación se enfatiza en una de estas tres áreas, dedicando cerca del 60% de los ítems a la evaluación del área en la que se enfoca la prueba.

La primera de estas evaluaciones se aplicó en el año 2000 y tuvo énfasis en el área de lectura; en esta evaluación participaron 32 países. La segunda prueba se aplicó en el año 2003, se concentró en las matemáticas y contó con la participación de 41 países. La tercera ocasión en que se evaluó estudiantes con la metodología de PISA fue en el año 2006, y concentró mayores esfuerzos en evaluar los conocimientos y habilidades en ciencias; ese año participaron 57 países. Con estas tres pruebas concluyó el primer ciclo completo de evaluaciones de PISA.

El segundo ciclo de pruebas se inició en el año 2009 con la evaluación que enfatiza lectura. Esta fue la primera evaluación en la que Costa Rica tomó parte dentro del programa de PISA. La segunda prueba en la que participó Costa Rica se realizó en el año 2012 y tuvo énfasis en las habilidades matemáticas, mientras que recientemente en mayo del año 2015 fue aplicada en nuestro país la prueba PISA 2015, cuyo énfasis son las habilidades en ciencias.

Además de la medición de habilidades y conocimientos de los estudiantes mediante una prueba escrita, PISA también realiza cuestionarios para recolectar información que permita analizar y contextualizar adecuadamente los resultados de las pruebas. Se aplican cuestionarios a los estudiantes para obtener información sobre aspectos del hogar y la familia. Asimismo, se aplican cuestionarios a los directores de cada centro educativo para obtener información específica sobre los aspectos propios de la infraestructura del centro educativo, cantidad y preparación del personal docente, entre otros aspectos.

Los valores del rendimiento de los estudiantes en las evaluaciones son considerados como variables latentes continuas, lo que significa que se tratan como cantidades que son medibles pero para las que no existe un instrumento de medida, ya que los ítems no son evaluados con puntuaciones específicas, sino que PISA define escalas a partir de las cuales se realizan las comparaciones. Así, para facilitar la interpretación de las notas asignadas a los estudiantes, las escalas en lectura, matemáticas y ciencias fueron diseñadas para tener una nota promedio de 500 puntos y una desviación estándar de 100 puntos para los países miembros de la OECD. Esto significa que cerca de

dos tercios de los estudiantes de países miembros de la OECD obtienen notas entre los 400 y 600 puntos (OECD, 2009). A estos valores del rendimiento de los estudiantes se les llama Valores Plausibles.

4.2.1.1. El Constructo de Alfabetización matemática

La evaluación PISA del año 2012 se enfocó en medir si los estudiantes son competentes desde el punto de vista matemático (o si adquirieron la Competencia Matemática), lo cual implica tener la capacidad individual para utilizar los conceptos aprendidos durante los procesos de aprendizaje en la resolución de problemas y en situaciones que se le presentan en otros contextos de su vida cotidiana. En esta evaluación el foco no se centra entonces en conocer cuáles contenidos del currículo han sido aprendidos por el alumno, sino cómo este los pone en práctica (Salas, 2012).

El marco conceptual bajo el cual se define la alfabetización matemática en PISA fue escrito bajo la guía de un Grupo de Expertos Matemáticas (Mathematics Expert Group, MEG) conformado por diez miembros, incluyendo matemáticos, educadores de matemáticas y expertos en evaluación, tecnología e investigación educativa de un grupo de países.

Adicionalmente, para asegurar una mayor robustez del marco conceptual, una versión preliminar del Marco Conceptual Matemático preparado por PISA fue circulado a más de 170 expertos matemáticos en más de 40 países, para recibir retroalimentación por parte de estos (OECD, 2013).

Según PISA (OECD, 2013), las matemáticas son una herramienta crítica para las personas jóvenes ya que estos enfrentan asuntos y retos relacionados con esta en los ámbitos personal, ocupacional y social. Por esto, es importante tener un entendimiento del grado en el que las personas jóvenes están adecuadamente preparados para aplicar las matemáticas, para comprender asuntos importantes y resolver problemas significativos en su diario vivir. Una evaluación a la edad de 15 años provee una indicación temprana de como los individuos podrían responder en un futuro al diverso conjunto de situaciones que se encontrarán en sus vidas y que involucre el uso de las matemáticas.

La concepción de alfabetización matemática apoya la importancia de que los estudiantes desarrollen una fuerte comprensión de los conceptos de la matemática pura y los beneficios de estar comprometidos con la exploración del mundo abstracto de las matemáticas. El constructo de alfabetización matemática, según la definición de PISA, hace un remarcado énfasis en la necesidad de desarrollar la capacidad de los estudiantes de utilizar las matemáticas de acuerdo con el contexto, para lo cual es importante que los estudiantes tengan ricas experiencias en sus clases de matemáticas (OECD, 2013).

El marco conceptual de la alfabetización matemática en PISA distingue entre grandes categorías de contenidos que son importantes para la alfabetización en general, y los contenidos específicos de temas que son apropiados para los estudiantes de 15 años. La alfabetización matemática no es un atributo que un individuo tiene o no tiene, sino que está en continuo crecimiento, aunque sí se reconoce que algunos individuos son más alfabetizados matemáticamente que otros (OECD, 2013).

Para los propósitos de PISA, la alfabetización matemática se define como sigue:

“La alfabetización matemática es la capacidad de un individuo para formular, emplear e interpretar las matemáticas en una variedad de contextos. Incluye razonar matemáticamente y utilizar conceptos matemáticos, procedimientos, datos y herramientas para describir, explicar y predecir fenómenos. Ayuda a los individuos a reconocer el rol que las matemáticas juegan en el mundo y para tomar las decisiones mejor fundamentadas que requiere un ciudadano constructivo, comprometido y reflexivo” (OECD, 2013).

Los conceptos de “formular”, “emplear” e “interpretar” presentes en la definición anterior apuntan a los tres procesos en los cuales los estudiantes participarán en su ciclo de vida.

Formular matemáticas involucra identificar oportunidades para aplicar y utilizar las matemáticas, entendiendo que las matemáticas pueden aplicarse para comprender o resolver un problema o reto particular. Incluye tener la capacidad de tomar una situación como se presente, y transformarla en una forma susceptible de tratamiento matemático, proveyendo las estructuras matemáticas y las representaciones, identificando variables y haciendo supuestos simplificadores para ayudar a resolver el problema.

Emplear matemáticas involucra aplicar el razonamiento matemático y utilizar los conceptos matemáticos, procedimientos, datos y herramientas para derivar una solución matemática. Incluye la realización de cálculos, la manipulación de expresiones y ecuaciones algebraicas u otros modelos matemáticos, el análisis de información en una forma matemática desde gráficos y diagramas matemáticos, desarrollar descripciones y explicaciones matemáticas, y el uso de herramientas matemáticas para resolver problemas.

Interpretar matemáticas involucra reflexionar sobre soluciones o resultados matemáticos e interpretarlos en el contexto de un problema o desafío. Incluye la evaluación de las soluciones matemáticas en relación con el contexto del problema y determinar si los resultados son razonables y tienen sentido para esa situación.

De acuerdo con el marco conceptual de PISA, la distribución de los ítems en la prueba de acuerdo con los procesos es la siguiente: 25% formulación matemática, 50% emplear conceptos matemáticas y 25% interpretación matemática.

4.2.1.2. Delimitación del dominio matemático

El marco conceptual de PISA define el dominio matemático para la encuesta de PISA y describe un enfoque para la evaluación de la alfabetización matemática de los jóvenes de 15 años de edad, es decir, PISA evalúa el grado de experticia en que los alumnos de 15 años de edad pueden manejar las matemáticas cuando se enfrentan a situaciones y problemas, la mayoría de las cuales se presentan en contextos del mundo real.

Para propósitos de la evaluación, la definición de alfabetización matemática de PISA 2012 puede ser analizada en términos de tres aspectos interrelacionados:

- Los *procesos matemáticos*, que describen lo que los individuos hacen para conectar el contexto del problema con las matemáticas y de esta manera resolver el problema, y las capacidades subyacentes a este proceso;
- El *contenido matemático* que está dirigido a utilizarse en los ítems de la evaluación; y
- Los *contextos* en los cuales los ítems evaluados se encuentran ubicados.

Para organizar el dominio matemático a efectos de evaluar la alfabetización matemática, fue definido un set de categorías de contenido que reflejan los fenómenos subyacentes a las matemáticas, que consiste en cuatro categorías utilizadas previamente en PISA. De acuerdo con PISA (2013), dichas categorías reúnen los requerimientos del desarrollo histórico que abarcan el dominio de las matemáticas e ilustran las grandes áreas de contenido que guían el desarrollo de los ítems en la evaluación del año 2012; estas son:

- Cambios y relaciones;
- Espacios y formas;
- Cantidades; y
- Datos e incertidumbre.

Si bien explicar estas categorías van más allá del alcance de este trabajo, una explicación detallada de estas puede encontrarse en OECD (2013). Sin embargo, es importante aclarar que dichas categorías no son excluyentes con respecto a los ítems, es decir, en la evaluación de PISA un ítem puede estar evaluando una, dos o más de estas categorías.

Según con el marco conceptual de PISA, la distribución de los ítems en la prueba de acuerdo con estas cuatro categorías se realiza de manera aproximadamente igual, de 25% del puntaje total de la evaluación en cada categoría.

Finalmente, PISA hace un gran esfuerzo por evaluar los niveles matemáticos que son apropiados para estudiantes de 15 años de edad, y que les ayudarán a convertirse en ciudadanos constructivos, comprometidos y reflexivos. Aunque la prueba PISA no pretende ser una evaluación de currículos específicos, sí busca reflejar el nivel de conocimiento de los conceptos matemáticos los estudiantes han aprendido a lo largo de su proceso educativo (OECD, 2013).

Estos conceptos no pueden ser asociados, en una relación uno a uno, con respecto a las cuatro categorías anteriores, pero se refieren básicamente a las siguientes:

- Funciones: el concepto de funciones, con énfasis en funciones lineales pero no limitado a estas.
- Expresiones algebraicas.
- Sistemas de coordenadas: representación y descripción de datos, posiciones y relaciones.
- Relaciones dentro y entre objetos geométricos en dos y tres dimensiones.
- Medición: cuantificación de características de objetos y espacios, como medidas de ángulos, distancias, tamaños, circunferencias, entre otros.
- Números y unidades: incluyendo propiedades de números enteros y racionales, los aspectos relevantes de los números irracionales, así como las cantidades y unidades referidas a fenómenos tales como el tiempo, el dinero, peso, temperatura, distancia, área y volumen.

- Operaciones aritméticas.
- Porcentajes, ratios y proporciones.
- Principios de conteo: combinaciones y permutaciones simples.
- Recolección, representación e interpretación de datos.
- Variabilidad de datos: conceptos de variabilidad, distribución y tendencia central.
- Muestra y muestreo: incluyendo inferencias simples basadas en las propiedades de las muestras.
- Probabilidad: noción de eventos aleatorios, variación aleatoria, frecuencia de eventos y aspectos básicos del concepto de probabilidad.

4.2.1.3. Escalas de competencia

PISA no sólo reporta el rendimiento de los estudiantes como un puntaje general, sino también en términos del contenido evaluado, ubicando a los estudiantes en una escala que describe lo que los estudiantes saben y pueden hacer.

A estas escalas se les llama “escalas de competencia” en lugar de “escalas de desempeño” ya que reportan lo que los estudiantes saben y pueden hacer a ciertos niveles en lugar de lo que hicieron los estudiantes en la evaluación (una única ocasión). Se establece así ya que PISA está interesado en obtener el nivel de competencia general del estudiante, y no el rendimiento mostrado en una prueba específica.

El diseño de la evaluación de la prueba de PISA hace esto posible al utilizar técnicas modernas de modelación de respuestas a los ítems para estimar, simultáneamente, la habilidad de todos los estudiantes que fueron evaluados y la dificultad de todos los ítems de la evaluación. La metodología de PISA se basa en muestras de estudiantes y de ítems para efectuar estimaciones poblacionales: una muestra de estudiantes de 15 años de edad es seleccionada para representar a toda la población de 15 años del país, y una muestra de ítems de un gran conjunto de estos es administrada a cada estudiante.

La relación entre la habilidad del estudiante y la dificultad de los ítems es probabilística, lo que quiere decir que existe cierta probabilidad de que un estudiante particular pueda completar correctamente algún ítem. Con esta información, los expertos en PISA definieron seis niveles de competencia en matemática, que van desde el nivel 1, que indica que el alumno posee la capacidad de resolver problemas y ejercicios matemáticos rutinarios, hasta el nivel 6, donde el alumno es capaz de resolver problemas que requieren de razonamiento y habilidades matemáticas avanzadas.

PISA considera al nivel 2 como el nivel básico de competencias, en el que los estudiantes comienzan a demostrar las habilidades en lectura que les permitirá participar eficazmente en la vida productiva. El siguiente cuadro resume la descripción de las competencias deseadas en cada nivel:

Tabla 1 Descripción de las escalas de competencia en la evaluación matemática de PISA 2012

Nivel	Descripción
6	En este nivel el estudiante puede contextualizar, generalizar y utilizar la información basándose en sus investigaciones o en la modelización de una situación o problema complejo. El estudiante puede relacionar la información proveniente de diferentes fuentes o representaciones. El estudiante debe ser capaz de realizar razonamientos matemáticos avanzados, puede aplicar los conocimientos, la simbología y la matemática formal para entender y explicar nuevas situaciones y generar nuevas estrategias para dar respuestas a las mismas. En este nivel el estudiante puede expresarse en forma precisa, reflexiona sobre sus hallazgos e interpreta y reflexiona apropiadamente.
5	En este nivel los estudiantes pueden desarrollar y trabajar con modelos de situaciones complejas, además identificar las limitaciones de los mismos y especificar los supuestos considerados. Pueden seleccionar, comparar y evaluar adecuadas estrategias para la resolución de problemas y aplicarlas para resolver situaciones complejas relacionadas con los modelos que se plantean. Los estudiantes de este nivel pueden seleccionar estrategias apropiadas y utilizar en forma adecuada las habilidades de pensamiento y razonamiento que poseen. Establecen relaciones entre las representaciones, la simbología y las caracterizaciones formales. Por último son capaces de reflexionar sobre sus decisiones y están en capacidad de formular y comunicar sus interpretaciones y los razonamientos realizados para obtener sus respuestas.
4	En este nivel el estudiante puede trabajar eficazmente con modelos explícitos de situaciones complejas concretas, las mismas pueden implicar ciertas restricciones o suposiciones. El estudiante puede seleccionar e integrar diferentes representaciones, incluyendo representaciones simbólicas, vinculándolas directamente con situaciones del mundo real. Los estudiantes de este nivel pueden utilizar las habilidades desarrolladas previamente y son capaces de razonar en forma competente. El estudiante puede elaborar y comunicar explicaciones bien fundadas y basadas en las interpretaciones de las acciones realizadas para resolver un determinado problema.
3	En este nivel los estudiantes pueden ejecutar procedimientos claramente descritos, incluyendo aquellos que requieren la toma de decisiones secuenciales. Pueden seleccionar y aplicar estrategias sencillas de resolución de problemas. Los estudiantes de este nivel pueden interpretar y utilizar representaciones tomadas de diferentes fuentes de información y razonar directamente sobre ellas. Pueden desarrollar un reporte corto donde comunicando las interpretaciones y los razonamientos realizados y los resultados obtenidos.
2	En este nivel los estudiantes pueden interpretar y reconocer situaciones en contextos que requieren únicamente de la inferencia directa. Pueden extraer la información relevante a partir de una sola fuente y pueden utilizar un único modo de representación. Los estudiantes de este nivel pueden emplear algoritmos básicos, fórmulas y procedimientos convencionales. En este nivel el estudiante realiza razonamientos directos y hace interpretaciones literales de los resultados.
1	Los estudiantes de este nivel pueden responder a las preguntas que involucren contextos familiares, donde toda la información relevante está presente y las preguntas están claramente definidas. Estos estudiantes son capaces de identificar la información para llevar a cabo los procedimientos rutinarios de acuerdo con instrucciones directas, en situaciones explícitas. Pueden realizar acciones obvias y continuar inmediatamente con determinados estímulos.
Debajo del nivel 1	Los estudiantes en este nivel no tienen éxito en las tareas matemáticas más básicas medidas por PISA. Su patrón de respuestas hace prever que no serían capaces de resolver la mitad de las tareas en una prueba elaborada exclusivamente con preguntas del nivel 1. Estos estudiantes tienen serias dificultades para utilizar la matemática en cualquier contexto.

Fuente: Traducción tomada de Montero y otros, 2012.

4.2.2. Marco muestral

La población objetivo en cada país participante en la prueba PISA 2012 consistió de los estudiantes con edades entre los 15 años y tres meses completos y 16 años y dos meses completos de edad, que asisten a instituciones educativas oficiales en séptimo grado o un grado superior. Por esta razón, la unidad de análisis es el estudiante.

Como la prueba PISA se trata de una evaluación internacional, que debe ser aplicada en cada país participante, se requieren Administradores de Proyectos Nacionales (APN). En Costa Rica, el equipo de APN fue constituido en la Dirección de Gestión y Evaluación de la Calidad del Ministerio de Educación Pública.

A cada APN se le pidió definir, explicar y justificar su propia población objetivo, tratando en la medida de lo posible de que no se desvíe mucho de la definición de población dada por PISA. Con

base en esta definición, se le pidió a cada APN construir su marco muestral. Este marco debía contener todos los centros educativos que tuvieran estudiantes de 15 años y estuvieran en séptimo grado o superior, incluyendo centros educativos que podrían ser excluidos del marco posteriormente.

La calidad del marco muestral afecta los resultados de la encuesta directamente a través de las probabilidades de selección de los colegios y por lo tanto sus pesos, por lo que a cada APN se le instó a construir su marco de una manera lo más precisa posible. Además, les fue suministrado un manual para la preparación del marco muestral de colegios, en el cuál le fue indicado a cada APN las principales líneas a seguir para la construcción o depuración del marco muestral, incluyendo el envío de la población objetivo y las exclusiones, el envío del marco muestral y de los centros educativos excluidos (exclusiones a nivel de centro educativo y exclusiones dentro de cada centro educativo).

En el caso de Costa Rica, para el *marco muestral* se utilizó la información de los datos de colegios que dispone el Ministerio de Educación Pública. Este consiste en una lista de todos los colegios del país que incluye el nombre del colegio, su ubicación (provincia, cantón, distrito, dirección regional), zona (urbana o rural), rama (académico o técnico, diurno o nocturno), datos del director del colegio y la matrícula total del centro educativo.

Esta información la prepara el Departamento de Análisis Estadístico del MEP a través de la aplicación del Censo Escolar, que consiste en la aplicación de cuestionarios a cada centro educativo con el objetivo de obtener información de su matrícula, con algunas desagregaciones. Este censo se aplica tres veces al año (matrícula inicial, matrícula a medio periodo y matrícula final). El marco muestral de PISA 2012 utilizó datos del Primer Censo Escolar del año 2011 (matrícula inicial).

El APN (la Dirección de Gestión y Evaluación de la Calidad el MEP) solicitó al Departamento de Análisis Estadístico esta información, pero con datos adicionales, suministrados a través de otras bases de datos, que incluyeran los datos de matrícula de cada colegio por edad, por año cursado y por sexo, para así poder depurar el marco muestral.

El principal problema de este marco muestral es su rezago de información. Además, como la prueba se aplicó a estudiantes de 15 años en el 2012, los estudiantes utilizados fueron los que tenían 14 años en el 2011, y que se constituyen en los que potencialmente tendrían 15 años en el 2012.

El marco muestral estuvo constituido por más de 900 colegios.

El muestreo se realizó en dos etapas. Las unidades muestrales en la primera etapa consisten en los centros educativos (colegios) que tengan estudiantes de 15 años de edad. Las unidades muestrales en la segunda etapa fueron los estudiantes de los colegios muestreados. Una vez que los colegios fueron seleccionados para estar en la muestra, una lista completa de los estudiantes de 15 años fue preparada.

4.2.3. Diseño muestral

En nuestro país, y siguiendo los protocolos generales de PISA, se tomó una *muestra aleatoria en dos etapas con probabilidad proporcional al tamaño, estratificada de conglomerados* que estuvo

constituida de 193 instituciones, de ellas 167 eran públicas y 26 eran privadas o subvencionadas, 158 académicas y 35 técnicas, 124 urbanas y 69 rurales.

Previo al muestreo de centros educativos, los colegios fueron asignados a grupos mutuamente excluyentes basados en características de los colegios llamadas estratos explícitos, formados para mejorar la precisión de los estimadores. En Costa Rica la estratificación utilizada fue *público/privado, urbano / rural*.

Estos colegios fueron seleccionados con *probabilidad proporcional al tamaño (PPT)*. En cada uno de los colegios seleccionados se eligieron al azar, partir del listado general de todos los jóvenes de 15 años, una muestra de 30 estudiantes, independientemente del nivel en que estuvieran matriculados.

Como los colegios fueron muestreados con probabilidad proporcional al tamaño, los que tenían un tamaño pequeño (menor a 30) fueron incluidos con un tamaño de 30, haciendo que todos ellos tuvieran la misma probabilidad de ser seleccionados en la primera etapa.

La cobertura bruta de jóvenes de 15 años del país (estén o no en el sistema educativo) fue del 49,6%. Esto quiere decir que la muestra solo es representativa de un 50% de todos los jóvenes de 15 años²¹. El nivel de exclusión (porcentaje complementario) de 50,4% no solo se explica a partir de la población de 15 años que no está cubierta por el sistema educativo, sino que también incluye el considerable número de jóvenes que, al tener adecuaciones curriculares, no formaron parte del marco muestral de PISA en esa oportunidad, así como los jóvenes que están en el sistema de educación abierta.

Así, la muestra de 4602 estudiantes que representan a 40386 estudiantes de la población objetivo, se distribuye de la siguiente manera:

Tabla 2 Estadísticas descriptivas de la muestra de datos de PISA para Costa Rica según categoría. 2012

<u>Categoría</u>	<u>Porcentaje</u>		<u>Promedio en matemáticas</u>	
Género				
Mujeres	53,1	(.72)	395,9	(3.06)
Hombres	46,9	(.72)	419,5	(3.56)
Sector educativo al que asiste				
Público	84,6	(1,4)	395,8	(3.06)
Privado	15,4	(1,4)	470,1	(3.56)
Año de Colegio que cursa				
Séptimo	7,4	(.92)	339,6	(4.66)
Octavo	13,7	(.88)	362,8	(3.20)
Noveno	39,6	(1.27)	405,2	(2.86)
Décimo	39,1	(1.77)	436,7	(3.50)
Undécimo	0,2	(.08)		

Fuente: Elaboración propia con datos de PISA 2012

²¹ En la evaluación PISA del año 2015 este porcentaje mejoró hasta el 63%.

4.2.4. Imputación de datos faltantes

Una particularidad de esta prueba es que por primera vez, en el año 2012 fue aplicado un diseño con rotación de cuestionario. Este tipo de diseño es utilizado cuando se desea indagar mucha más información de relevancia para la toma de decisiones, en este caso de política educativa, pero sabiendo que existe un riesgo potencial para la aparición de no-respuestas por parte de los estudiantes al tener que contestar un cuestionario muy extenso.

Así, para lograr obtener toda la información necesaria del cuestionario, y prevenir un incremento en los datos faltantes, en la evaluación del 2012 fue introducido el diseño con rotación de cuestionario, únicamente en el cuestionario de contexto de los estudiantes²².

Esta rotación consiste en distribuir todos los paquetes temáticos del universo total de preguntas de dicho cuestionario en distintos formularios con varias combinaciones de estos paquetes temáticos en cada uno, por lo que en cada formulario se indaga un número menor de preguntas, pero al tomarlos todos juntos se garantiza la respuesta por parte de cada país de todos los items de la batería completa de preguntas.

Fueron formados tres formularios para la evaluación del 2012, donde cada formulario indaga dos terceras partes de la batería completa de preguntas, aunque todos incluyen una parte común de 25 items sobre la información básica de los estudiantes (como género, grado cursado e información sobre su hogar y su familia). De esta manera, se logra obtener un 33% más de información.

Estos formularios son rotados aleatoriamente entre todos los estudiantes de cada centro educativo, lo que garantiza que el proceso de datos omitidos se genera en forma completamente aleatoria (Missing completely at Random, MCAR), ya que la distribución de los valores observados no depende del patrón de comportamiento de los registros sin información.

Esta rotación aleatoria de los cuestionarios permite realizar el supuesto de distribución de datos faltantes como MCAR o como MAR, que es requerido para la aplicación de distintas técnicas de imputación sobre los valores faltantes.

Kaplan y Su (2015) analizaron los efectos de la aplicación de este diseño con rotación sobre las estimaciones posteriores, comparando varios métodos de imputación basados en la metodología de ecuaciones en cadena. Este enfoque utiliza un modelo de regresión univariado consistente con la escala de la variable con datos perdidos para proveer valores predichos dados los valores observados.

Una vez que una variable de interés se completa con los valores predichos, esa variable junto con las variables que tienen datos completos es usada en una secuencia para la siguiente variable que

²² Debido a que en la evaluación del 2012 fue introducido por primera vez en PISA el diseño con rotación de cuestionario, y los resultados de uso debían ser revisados y validados en etapas posteriores a la realización de la prueba para poder concluir sobre su robustez, se decidió no implementar este diseño en la evaluación del año 2015. No obstante, ya en el año 2015 se encontró evidencia de robustez en las estimaciones con este diseño (Kaplan y Su, 2015).

presenta valores faltantes. Una vez que la secuencia se completa para todas las variables con valores faltantes, la distribución posterior de los parámetros de regresión se obtiene vía muestreo de Gibbs y el proceso comienza de nuevo. Este algoritmo puede “correr” estas secuencias simultáneamente “m” número de veces obteniendo “m” set de datos imputados.

Kaplan y Su (2015) comparan este método bajo tres formas distintas de la generación de valores predichos: *predictive mean matching*, *Bayesian regression imputation* y *proportional odds logistic regression*, y encuentran que la técnica de *predictive mean matching* (PMM) es la más precisa y que el supuesto MCAR se cumple, ya que es posible conservar la distribución marginal de los valores plausibles y la estructura de covarianzas de los datos después de la imputación.

De esta manera, en esta tesis fue aplicada la técnica de imputación de datos faltantes *predictive mean matching* bajo el supuesto de que los datos faltantes se distribuyen completamente aleatorios. La imputación se hará con el paquete de R “mice”: Multivariate Imputation by Chained Equations (van Buuren y Groothuis-Oudshoorn (2011))²³.

4.2.5. Estimación del modelo

Los cuestionarios de contexto de la evaluación PISA 2012 incluyen una gran cantidad de ítems sobre las características de los estudiantes, extracción social de la familia, percepciones de los estudiantes, características de los centros educativos y la percepción de los directores de los centros educativos.

Algunos de los ítems fueron diseñados para ser utilizados en análisis como ítems singulares (por ejemplo el género). Sin embargo, la mayoría de los ítems de los cuestionarios fueron diseñados para ser combinados de cierta manera con el objetivo de medir constructos latentes que no pueden ser observados directamente (OECD, 2013).

Para estos ítems, fueron aplicados procedimientos de transformación o escalamiento para construir índices con sentido teórico y práctico. El alfa de Cronbach fue utilizado para evaluar la consistencia interna de cada índice dentro de los países y para compararla entre países.

A continuación se explican los constructos que fueron empleados y los indicadores que los conforman (entre paréntesis se muestra su codificación en las bases de datos original de PISA):

4.2.5.1. Constructos definidos para el nivel dentro de grupos

- **Constructo 1: Extracción socioeconómica (ExctSoci):** Es un constructo que intenta medir las condiciones sociales, económicas y familiares en las cuales se desarrolla el estudiante. Está conformado de cuatro variables observadas:

²³ Este fue el paquete que fue utilizado en Kaplan y Su (2015) para la imputación mediante predictive mean matching.

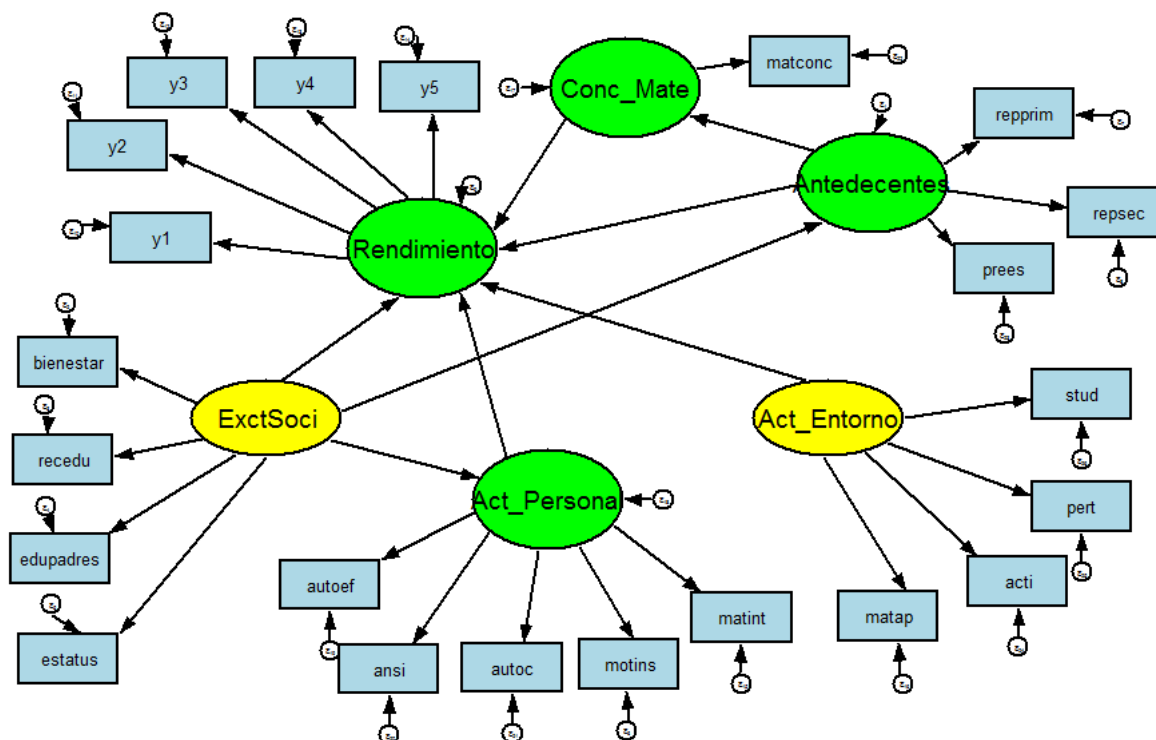
- **Posesiones de bienestar familiar** (bienestar): un índice que mide el bienestar familiar a través de algunos bienes presentes en el hogar, ya que se considera que medir el bienestar a través de estos bienes es una forma más estable que a través del ingreso. El estudiante debía responder si en su hogar había o no lo siguiente:
 - a. Una habitación solo para usted.
 - b. Conexión a internet.
 - c. Reproductor de DVD
 - d. Teléfonos celulares.
 - e. Televisores.
 - f. Computadoras.
 - g. Automóvil.
 - h. Habitación con baño privado.
 - i. Tres ítems adicionales pero específicos de cada país.
 - **Recursos educativos en el hogar** (rededu): Similar al anterior, pero midiendo artículos relevantes para el estudio en el hogar. Incluye lo siguiente:
 - a. Un escritorio o mesa para estudiar.
 - b. Un lugar tranquilo para estudiar.
 - c. Una computadora que puede usar para sus labores escolares.
 - d. Programas educativos para la computadora.
 - e. Libros de consulta para sus tareas escolares.
 - f. Libros de referencia técnica.
 - g. Un diccionario.
 - **Nivel educativo más alto de los padres** (edupadres): es medido mediante años de educación de acuerdo con el ISCED (International Standard Classification of Education).
 - **Estatus ocupacional más alto de los padres** (estatus): es medido mediante el Índice ISEI (International Socio-Economic Index of Occupational Status).
- **Constructo 2: Antecedentes educativos (Antecedentes):** Mide algunas características propias del estudiante relacionadas con su pasado educativo. Está conformado de tres variables dicotómicas:
- No repitió años en primaria (repprim).
 - No repitió años en secundaria (repsec)
 - Asistió a la educación preescolar por un año o más (prees).
- **Constructo 3: Actitud personal hacia las matemáticas (Act Personal):** Es un constructo que mide el grado en el que el estudiante se encuentra comprometido con la materia, disfruta del estudio de las matemáticas y su ansiedad a la materia. Está conformado de cinco variables:
- **Autoeficacia en matemáticas** (autoef): Ocho ítems fueron utilizados para medir este constructo, que se relaciona con el nivel de eficacia propia de cada estudiante para la resolución de problemas matemáticos. Dado cada ítem, el estudiante debía responder que tan seguro se siente al realizar las tareas matemáticas, en una escala que va de Muy seguro, Seguro, No muy seguro a Nada seguro. Dos ejemplos de estos ítems son los siguientes:
 - a. Calcular cuánto más barato vale un televisor con un 30% de descuento,
 - b. Resolver una ecuación como esta: $2(x + 3) = (x + 3)(x - 3)$

- **Ansiedad en matemáticas (ansi):** Cinco ítems fueron utilizados para medir este constructo, que se relaciona con el nivel de ansiedad con el que el estudiante aborda situaciones que involucren el uso de matemáticas. La escala de respuesta es: Totalmente de acuerdo, de acuerdo, en desacuerdo, totalmente en desacuerdo. Dos ejemplos de estos ítems son los siguientes:
 - a. Me pongo muy nervioso cuando me enfrento a un problema matemático.
 - b. Me preocupa sacar bajas notas en matemática.
 - **Auto concepto en matemáticas (autoc):** Cuatro ítems fueron utilizados para medir este constructo, en el que el estudiante se auto califica en cuanto a sus capacidades matemáticas. En este caso, la escala de respuestas se basan en qué tan probable es que el estudiante tenga estos pensamientos o sentimientos ante algunas situaciones. Dos ejemplos de estos ítems son los siguientes:
 - a. No soy bueno para resolver problemas de matemática.
 - b. A veces los temas de la lección son demasiado difíciles.
 - **Motivación instrumental hacia las matemáticas (motins):** Cuatro ítems fueron utilizados para medir este constructo, que se relaciona con la importancia que tiene para el estudiante el estudio de las matemáticas. La escala de respuesta es: Totalmente de acuerdo, de acuerdo, en desacuerdo, totalmente en desacuerdo. Dos ejemplos de estos ítems son los siguientes:
 - a. Hacer un esfuerzo en matemáticas vale la pena porque esto me ayudará en el trabajo que quiero conseguir en el futuro.
 - b. Matemáticas es una materia importante para mí porque la necesito para lo que quiero estudiar en el futuro.
 - **Interés en las matemáticas (matint):** Cuatro ítems fueron utilizados para medir este constructo, que se relaciona el interés propio y el gusto que tiene el estudiante hacia esta materia. La escala de respuesta es: Totalmente de acuerdo, de acuerdo, en desacuerdo, totalmente en desacuerdo. Dos ejemplos de estos ítems son los siguientes:
 - a. Espero con ansias mis lecciones de matemáticas.
 - b. Practico la matemática porque me gusta.
- **Constructo 4: Actitud hacia el entorno (Act Entorno):** Es un constructo que mide el grado en el que el estudiante se encuentra cómodo, satisfecho o feliz con el entorno que lo rodea en el centro educativo, incluyendo su relación con los docentes. Está conformado de cuatro variables:
- **Apoyo del profesor de matemáticas (matap):** Cuatro ítems fueron utilizados para medir este constructo. La escala de respuesta es: Totalmente de acuerdo, de acuerdo, en desacuerdo, totalmente en desacuerdo. Dos ejemplos de estos ítems son los siguientes:
 - a. Mi profesor nos da ayuda extra cuando la necesitamos.
 - b. Mi profesor da oportunidad a los estudiantes para que expresen sus opiniones.
 - **Relaciones estudiante profesor (stud):** Cuatro ítems fueron utilizados las relaciones entre el estudiante con los profesores de su centro educativo. La escala de respuesta es: Totalmente de acuerdo, de acuerdo, en desacuerdo, totalmente en desacuerdo. Dos ejemplos de estos ítems son los siguientes:
 - a. La mayoría de los profesores se interesan en el bienestar de los estudiantes.

- b. Los estudiantes se llevan bien con la mayoría de los profesores.
 - **Actitud hacia el centro educativo (acti):** Cuatro ítems fueron utilizados para medir este constructo, que se relaciona con la actitud del estudiante hacia el centro de enseñanza al que asiste. La escala de respuesta es: Totalmente de acuerdo, de acuerdo, en desacuerdo, totalmente en desacuerdo. Dos ejemplos de estos ítems son los siguientes:
 - a. El colegio ha sido una pérdida de tiempo.
 - b. El colegio me ha enseñado cosas que pueden ser útiles en un trabajo.
 - **Sentido de pertenencia al centro educativo (pert):** Nueve ítems fueron utilizados para medir este constructo, que se relaciona con el sentido propio del estudiante de pertenencia a su colegio. La escala de respuesta es: Totalmente de acuerdo, de acuerdo, en desacuerdo, totalmente en desacuerdo. Dos ejemplos de estos ítems son los siguientes:
 - a. El colegio me siento incómodo y fuera de lugar.
 - b. En el colegio hago amigos fácilmente.
- **Constructo 5: Familiaridad con conceptos matemáticos (matconc):** Es un constructo que mide el nivel de familiaridad del estudiante con conceptos matemáticos. Es un constructo previamente calculado por PISA (FAMCON), por lo que sería incluido al modelo como una única variable. Sin embargo, el índice está formado por trece ítems que miden la familiaridad auto percibida del estudiante a algunos conceptos matemáticos como la función exponencial, función cuadrática, divisor, número racional, coseno, y otros. La escala de respuesta es: Nunca oí hablar de eso, Oí hablar de eso una o dos veces, He oído hablar de eso algunas veces, He oído hablar de eso frecuentemente, Lo conozco bien y entiendo el concepto.
- **Constructo 6: Rendimiento en matemáticas (Rendimiento):** Es un constructo que evalúa en una escala latente la competencia matemática de los estudiantes. Está conformado por los cinco valores plausibles de la evaluación de matemáticas de la prueba PISA 2012 (y1-y5 en la figura siguiente).

La siguiente figura muestra la forma en que se relacionan estructuralmente los constructos recién explicados. Interesa particularmente las relaciones del estrato socioeconómico con el rendimiento educativo, tanto su relación directa, como sus relaciones indirectas a través de los antecedentes educativos y la actitud del estudiante:

Figura 5 Relación entre factores del modelo de ecuaciones estructurales a ser estimado con datos de la evaluación PISA 2012



4.2.5.2. Constructos definidos para el nivel entre grupos

Constructo 1: Calidad de recursos educativos: Es un constructo compuesto de tres indicadores mediante el cual el director del centro educativo evalúa la capacidad de los principales recursos educativos con los que cuenta el centro educativo.

- **Escasez de personal docente (TCHORT):** Es un constructo compuesto de cuatro indicadores mediante el cual el director del centro educativo indicaba si la capacidad de su escuela para proporcionar educación se ve obstaculizada por falta de personal calificado en ciencias, matemáticas, idiomas u otras materias.
- **Calidad de los recursos educativos del centro educativo (SCMATEDU):** El índice de los recursos educativos de la escuela se calculó sobre la base de seis ítems que miden las percepciones de los directores de la escuela sobre los posibles factores que obstaculizan la instrucción en la escuela. Estos incluyen carencia de laboratorios equipados adecuadamente, de materiales educativos, computadoras para fines pedagógicos, conexión a internet, software educativo y libros.
- **Calidad de la infraestructura física (SCMATBUI):** Mide la percepción del centro educativo de factores que potencialmente podrían afectar el proceso educativo relacionado con infraestructura. Incluye edificios escolares y terrenos, sistemas de iluminación y calefacción o enfriamiento, y espacios educativos como aulas y talleres.

Constructo 2: Clima escolar: Es un constructo compuesto por dos sub constructos para medir el clima escolar del centro educativo en sus dos ejes principales: estudiantes y docentes.

- **Factores relacionados con los estudiantes que afectan el clima escolar (STUDCLIM):** Incluye ausentismo escolar, estudiantes escapando de clases, llegadas tardías de los estudiantes, la no asistencia a eventos obligatorios, falta de respeto a los docentes, uso de alcohol o drogas ilegales y bullying.
- **Factores relacionados con los docentes que afectan el clima escolar (TEACCLIM):** Este índice incluye aspectos relacionados con la relación de docentes hacia estudiantes que podrían afectar el proceso educativo, como pobres relaciones entre estudiantes y docentes, clases compuestas por estudiantes con distintas habilidades, bajas expectativas de los docentes para con los estudiantes, docentes que no atienden las necesidades educativas individuales, ausentismo docente, resistencia al cambio de parte del personal docente, entre otros..

Constructo 3: Calidad docente: Es un constructo compuesto de dos indicadores mediante el cual el director del centro educativo indica su percepción respecto a dos variables clave del desenvolvimiento de sus docentes:

- **Moral docente (TCMORALE):** Es un constructo compuesto de cuatro indicadores mediante el cual el director del centro educativo indica su percepción respecto de la moral de los docentes, el entusiasmo de los mismos, el orgullo de los docentes de trabajar en la institución y si estos valoran los resultados académicos sobre otros factores.
- **Focalización docente (TCFOCST):** Mediante este índice se indica si los docentes de matemáticas están interesados en probar nuevos métodos y prácticas de enseñanza en sus clases; además, si existe o no consenso entre los docentes de matemáticas sobre la necesidad de adaptar los estándares académicos al nivel de los estudiantes y sus necesidades.

La relación que fue modelada entre estos tres constructos en el nivel superior es de asociación directa con el rendimiento académico.

5. Resultados

5.1. Resultados del estudio de simulación

Esta sección presenta los hallazgos para las simulaciones realizadas según el diseño que fue descrito en la sección 4.1. Se analizan los resultados obtenidos al evaluar tres puntos:

1. La estabilidad de las estimaciones de los parámetros.
2. La cantidad de sesgo introducida en las estimaciones de los parámetros.
3. El ajuste y comparación de los modelos mediante el cómputo del *P-value* predictivo a posteriori.

Para la estabilidad de las estimaciones de los parámetros, fue examinada la posición de la distribución de las estimaciones producidas en las simulaciones y la raíz cuadrada del error cuadrático medio asociado a cada parámetro. Para la cantidad de sesgo en las estimaciones, fue calculada la diferencia entre el valor dado del parámetro y el valor estimado, de forma que se puedan apreciar las desviaciones -o no- del valor de la pendiente desde el cual se simulan los datos. Finalmente, el cambio en el ajuste de los modelos se evaluará mediante el cómputo de la probabilidad asociada a la prueba cociente de verosimilitudes del modelo nulo respecto del modelo irrestricto (*haciendo uso del P-value predictivo a posteriori*).

Es importante aclarar que no fueron explicados los resultados en términos de la convergencia de los distintos modelos debido a que todas las replicaciones requeridas convergieron para todas las condiciones de estimación Bayesiana. Sin embargo, la convergencia requirió más iteraciones para los modelos con las distribuciones a priori difusas y las muestras con menor número de observaciones.

5.1.1 Estabilidad de las estimaciones de los parámetros

Todos los modelos estimados comparten la estructura factorial presentada en la figura 1 (sección 4.1.1). Los parámetros de interés de esa estructura factorial no se limitan a las cargas factoriales, sino que incluyen las siguientes partes (entre paréntesis se muestra el número de parámetros estimados libremente en cada caso):

- Las cargas factoriales directas (4)
- Las cargas factoriales cruzadas (6)
- El parámetro estructural (1)
- Las varianzas residuales de las variables indicadoras (6)
- Las varianzas residual del factor 1 y la varianzas factorial del factor 2 (2)

En total, el modelo general contiene 19 parámetros de interés. La estrategia principal consiste en la comparación de resultados para los tres tipos de distribuciones a priori con las que fueron modeladas las cargas factoriales cruzadas (difusa, débilmente informativa, informativa). Además, resulta de interés comparar la estabilidad para cuatro tamaños de muestra distintos y con tres distintos valores de la fuerza del coeficiente de correlación intraclase.

La estrategia indicada en el párrafo anterior fue replicada para seis distintos escenarios de modelación de las distribuciones a priori del resto de parámetros a estimar (parámetros de regresión y de varianzas).

Finalmente, los 19 parámetros de interés fueron estimados tanto en el nivel dentro de grupos como en el nivel superior entre grupos, es decir, cada uno fue estimado en los dos niveles, por lo que en total de cada modelo podrían ser analizados 38 parámetros: 19 del nivel inferior y 19 del nivel superior.

Dada la elevada cantidad de modelos que fueron estimados y la gran cantidad de parámetros que pudieran ser analizados y reportados, y con el objetivo de reducir en la medida de lo posible la cantidad de información que debe ser visualizada, será presentado el resultado de los parámetros según las 5 agrupaciones enumeradas anteriormente, en lugar de presentar resultados para los 19 parámetros individuales. De esta forma, en el caso de las 4 cargas factoriales directas, no se mostrarán los resultados de estabilidad y sesgo de cada uno, sino que será presentado el resultado del parámetro con la menor estabilidad o con el mayor sesgo posible, como referencia para comentar sobre ese agrupamiento en particular.

La estrategia de presentación consiste en separar el análisis según el nivel de agrupamiento (nivel inferior y nivel superior) para los seis escenarios de modelación de las distribuciones a priori de todos los parámetros de regresión y de varianza. Así, cada gráfico incluye de forma agrupada, los resultados de los 12 modelos estimados bajo esas condiciones particulares (3 ICC x 4 tamaños de grupo) y cada sub gráfico compara visualmente las estimaciones condicionadas a esas características según el tipo de distribución a priori establecido para las cargas factoriales cruzadas (difusas, débilmente informativas e informativas). Es decir, cada gráfico muestra 36 modelos.

5.1.1.1 Las distribuciones de las estimaciones de los parámetros

Para la evaluación de la estabilidad primero se examinaron, en cada escenario, las distribuciones de las estimaciones de los parámetros mediante gráficos que muestran el intervalo de credibilidad de cada parámetro. Estos resultados se muestran de forma completa en los anexos 2 y 3.

5.1.1.1.1 Nivel inferior (dentro de grupos)

Los gráficos 1 al 6 del anexo 2 muestran que, manteniendo lo demás constante, el uso de distribuciones a priori informativas en las cargas factoriales cruzadas mejora la precisión con la que son estimados los parámetros de las cargas factoriales directas, ya que el intervalo de credibilidad al 95% de estas estimaciones se reduce consistentemente al aumentar la cantidad de información en las varianzas cruzadas.

Estas mejoras son incluso más pronunciadas cuando se toma en cuenta el tamaño de la muestra, ya que entre mayor número de datos menor es el intervalo de credibilidad del parámetro, especialmente en los casos con gran número de grupos. Finalmente, se aprecian algunas bondades al aumentar el coeficiente de correlación intraclase de 0.02 a 0.05, no así en el cambio a 0.40.

Además, entre menos informativas sean las distribuciones a priori del resto de parámetros de interés, también mayor amplitud tendrán los intervalos de credibilidad.

En cuanto a la estimación de las propias cargas factoriales cruzadas (gráficos 7 al 12), es claro que el uso de varianzas difusas en sus distribuciones a priori produce malos resultados tanto en términos de su precisión como en su sesgo, lo cual se mantiene para las distintas combinaciones de escenarios, empeorando sus resultados conforme lo esperado: menor tamaño de muestra, menor correlación intraclase y menor información en la distribución a priori del resto de parámetros de interés.

La estimación del parámetro estructural del modelo simulado muestra un comportamiento similar al ya comentado, pero con el agravante de que el uso de informaciones a priori difusas en las cargas factoriales cruzadas producen muy pobres resultados y, en general, el parámetro estructural no logra ser recuperado adecuadamente en estos casos (gráficos 13 a 18).

En el caso de las varianzas residuales, los intervalos de credibilidad mejoran según lo esperado dados los escenarios (a mayores tamaños de muestra, a mayor coeficiente de correlación intraclase), pero de entrada la varianza de sus estimaciones es mayor que la del resto de parámetros, exacerbando la dispersión en el caso del coeficiente de correlación intraclase más bajo (0.02).

Las varianzas factoriales, por su parte, presentan poca precisión en sus estimaciones (lo cual se comentará en mayor detalle en el análisis de la raíz cuadrada del error cuadrático medio), pero llama la atención que en este caso el patrón se invierte, y es en los casos con mayor tamaño de muestra en donde la dispersión se acentúa y produce grandes intervalos de credibilidad, particularmente bajo la condición de varianzas a priori difusas en las cargas factoriales cruzadas.

5.1.1.1.2 Nivel superior (entre grupos)

En el anexo 3 se muestran los 30 gráficos correspondientes a las estimaciones de los parámetros para el nivel superior dentro de la jerarquía del modelo (gráficos 31 al 60). Como es esperable de acuerdo con la teoría, el nivel de precisión y el sesgo de las estimaciones es más elevado en el nivel superior del modelo comparado con el nivel inferior, por lo cual el análisis de este nivel tendrá un peso mayor en la elaboración de las conclusiones con base en los resultados de las simulaciones.

A nivel general, la estimación Bayesiana con varianzas a priori informativas en las cargas factoriales cruzadas presenta resultados superiores a los de las demás condiciones de estimación en el nivel entre grupos. Sin embargo, sí existen variaciones en cuanto al nivel de sesgo dependiendo del tipo de parámetro.

Para el parámetro de cargas factoriales directas (cuyo valor poblacional es de 1) el sesgo de las estimaciones es pequeño, es decir, el intervalo de credibilidad se encuentra bastante centrado el nivel del valor poblacional, pero el nivel de precisión empeora bastante en comparación con el nivel dentro de grupos.

En el caso de los parámetros estimados en el escenario 1 (que incluye las distribuciones a priori más informativas para el resto de parámetros de varianza y regresión) el nivel de precisión es similar tanto para distintos tamaños de muestra como entre los distintos valores de la correlación intraclase. No obstante, a medida que disminuye la información en la varianza de la distribución a priori de estos otros parámetros distintos a las cargas factoriales cruzadas (escenarios 2 al 6), la

precisión empeora aún más, particularmente en el caso de distribuciones a priori poco informativas en las cargas factoriales cruzadas.

Es decir, para este parámetro, únicamente una distribución a priori informativa en las cargas factoriales cruzadas garantiza niveles aceptables de precisión de la estimación para cualquier escenario simulado.

La estimación de las cargas factoriales cruzadas no es de interés en sí misma, sin embargo, el anexo 3 muestra que la distribución a priori difusa genera resultados indeseables tanto en sesgo como en precisión.

En cuanto al parámetro estructural, el escenario 1 (que agrega un nivel de información superior a la estimación de parámetros) muestra estimaciones poco sesgadas y con aceptable nivel de precisión, e incluso este nivel de precisión es poco sensible a otras variaciones del modelo (prácticamente no cambia en los distintos tamaños de muestra, ni ante variaciones en el ICC), pero sí resulta sensible a la cantidad de información aportada a la varianza de la distribución a priori de las cargas factoriales cruzadas (produciendo mejores resultados a mayor información).

Contrario a lo que se observó en la estimación dentro de grupos, donde las varianzas residuales mostraron una alta dispersión bajo los escenarios relacionados al coeficiente de correlación intraclase de 0.02, en el caso del modelo entre grupos tanto la precisión como el sesgo mostraron pobres resultados para los escenarios de un coeficiente de correlación intraclase de 0.40.

Las varianzas factoriales tuvieron el mismo comportamiento (peores resultados para el ICC=0.4), pero maximizado en el hecho de que, para distribuciones a priori difusas en las cargas factoriales cruzadas, el parámetro ni siquiera logró ser recuperado (es decir, el intervalo de credibilidad no incluye el valor poblacional)

En los casos anteriores (varianza residual y factorial) la estimación mediante distribución a priori muy informativa en la carga factorial cruzada produjo resultados similares al uso de información débilmente informativa.

5.1.1.2 Raíz cuadrada del error cuadrático medio en los parámetros

La Raíz Cuadrada del error cuadrático medio (RMSE, siglas en inglés) se muestra en las tablas 1 a la 6 para el nivel dentro de grupos, y tablas 7 a la 12 para el nivel dentro de grupos.

5.1.1.2.1 Nivel inferior (dentro de grupos)

El RMSE o la desviación estándar de los residuales es una medida de qué tan lejos están los puntos de datos de la línea de regresión; el RMSE es una medida de la dispersión de estos residuos. Este puede ser interpretado como una medida de la dispersión de las estimaciones con respecto a su valor poblacional.

En las tablas siguientes (tablas 1 a 6) se muestran los resultados de esta medida para cuatro tipos de parámetros: la carga factorial directa, la carga factorial cruzada, el parámetro estructural y la varianza residual (incluyendo la varianza factorial).

Al igual que con el caso de las distribuciones de las estimaciones, para presentar el RMSE se recurre a mostrar el valor más alto para cada tipo de parámetro dentro del mismo escenario. Además, se establece el umbral de 0.5 como medida arbitraria para evidenciar en cuáles casos se obtiene una pobre estimación del parámetro.

En el caso del nivel dentro de grupos, el RMSE mostró altos valores en los parámetros de carga factorial cruzada, parámetro estructural y varianza residual, pero únicamente cuando la distribución a priori de la carga factorial cruzada es difusa. Sólo en el escenario 6 las cargas factoriales directas también presentaron algunos inconvenientes con el valor del RMSE, especialmente para tamaños de muestra pequeños y con un ICC bajo (0.02).

En general, el RMSE cuando la distribución a priori de la carga factorial cruzada es débilmente informativo o informativo presenta un buen comportamiento.

Tabla 3 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 1)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.082	0.047	0.020	23.470	0.148	0.006	0.110	0.046	0.019	0.794	0.378	0.068
	G=30, n=30	0.062	0.033	0.016	22.007	0.150	0.005	0.109	0.044	0.012	0.768	0.352	0.029
	G=100, n=10	0.060	0.032	0.016	25.606	0.136	0.004	0.102	0.055	0.011	0.837	0.347	0.026
	G=100, n=30	0.043	0.016	0.010	23.500	0.171	0.003	0.088	0.052	0.007	1.218	0.391	0.018
ICC=0.05	G=30, n=10	0.068	0.026	0.014	30.219	0.147	0.005	0.114	0.045	0.012	0.811	0.381	0.029
	G=30, n=30	0.036	0.011	0.007	22.909	0.178	0.003	0.130	0.049	0.007	0.809	0.417	0.015
	G=100, n=10	0.042	0.012	0.007	29.671	0.179	0.003	0.115	0.049	0.006	0.956	0.418	0.012
	G=100, n=30	0.013	0.004	0.003	14.481	0.175	0.002	0.092	0.051	0.004	1.454	0.444	0.010
ICC=0.40	G=30, n=10	0.059	0.025	0.016	25.843	0.153	0.005	0.102	0.041	0.014	0.348	0.159	0.014
	G=30, n=30	0.035	0.011	0.008	20.798	0.175	0.004	0.118	0.051	0.008	0.532	0.175	0.008
	G=100, n=10	0.031	0.011	0.008	19.684	0.165	0.003	0.116	0.045	0.006	0.394	0.149	0.007
	G=100, n=30	0.012	0.004	0.003	12.374	0.155	0.002	0.088	0.046	0.004	0.333	0.142	0.004

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 4 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 2)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.215	0.119	0.046	23.253	0.210	0.007	0.799	0.151	0.039	0.760	0.420	0.076
	G=30, n=30	0.152	0.070	0.027	25.247	0.232	0.006	0.662	0.186	0.019	0.894	0.424	0.038
	G=100, n=10	0.158	0.062	0.026	25.761	0.220	0.005	0.816	0.189	0.018	0.775	0.433	0.038
	G=100, n=30	0.085	0.022	0.012	24.666	0.237	0.004	0.601	0.184	0.008	0.908	0.427	0.021
ICC=0.05	G=30, n=10	0.160	0.040	0.019	33.238	0.228	0.005	0.701	0.156	0.018	0.814	0.438	0.033
	G=30, n=30	0.071	0.013	0.008	29.471	0.242	0.004	0.557	0.184	0.009	0.942	0.477	0.018
	G=100, n=10	0.084	0.014	0.008	34.190	0.267	0.003	0.632	0.193	0.007	0.966	0.426	0.016
	G=100, n=30	0.024	0.004	0.003	20.762	0.232	0.002	0.317	0.198	0.004	0.866	0.336	0.011
ICC=0.40	G=30, n=10	0.165	0.040	0.020	34.780	0.237	0.006	0.683	0.167	0.019	0.364	0.175	0.016
	G=30, n=30	0.080	0.012	0.009	30.261	0.245	0.004	0.516	0.177	0.008	0.492	0.159	0.008
	G=100, n=10	0.080	0.012	0.007	31.712	0.251	0.003	0.584	0.190	0.007	0.425	0.157	0.007
	G=100, n=30	0.031	0.004	0.003	22.078	0.248	0.002	0.363	0.194	0.004	0.497	0.140	0.005

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 5 Error cuadrático medio de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 3)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.356	0.176	0.056	27.602	0.219	0.007	1.645	0.223	0.046	0.781	0.465	0.079
	G=30, n=30	0.261	0.082	0.031	30.903	0.253	0.006	1.338	0.305	0.022	0.982	0.453	0.041
	G=100, n=10	0.273	0.075	0.028	31.604	0.257	0.005	1.765	0.348	0.020	0.843	0.481	0.041
	G=100, n=30	0.145	0.024	0.014	30.465	0.269	0.004	1.146	0.355	0.010	1.388	0.410	0.024
ICC=0.05	G=30, n=10	0.252	0.049	0.022	42.050	0.246	0.006	1.403	0.275	0.019	0.836	0.437	0.034
	G=30, n=30	0.109	0.013	0.009	39.246	0.278	0.004	1.040	0.324	0.009	0.945	0.415	0.019
	G=100, n=10	0.106	0.014	0.008	38.629	0.300	0.004	1.332	0.367	0.008	0.851	0.427	0.016
	G=100, n=30	0.028	0.004	0.003	23.901	0.251	0.002	0.726	0.279	0.004	0.843	0.504	0.009
ICC=0.40	G=30, n=10	0.255	0.044	0.023	42.411	0.250	0.006	1.450	0.288	0.021	0.323	0.190	0.018
	G=30, n=30	0.122	0.014	0.009	43.946	0.270	0.004	1.102	0.300	0.009	0.403	0.164	0.009
	G=100, n=10	0.117	0.012	0.008	41.348	0.267	0.004	1.207	0.361	0.008	0.330	0.175	0.007
	G=100, n=30	0.030	0.004	0.003	24.124	0.258	0.002	0.606	0.333	0.004	0.326	0.159	0.005

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 6 **Error cuadrático medio** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 4)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.285	0.128	0.042	31.092	0.212	0.007	0.823	0.152	0.036	0.780	0.438	0.075
	G=30, n=30	0.183	0.063	0.028	32.075	0.235	0.006	0.770	0.172	0.019	0.876	0.449	0.040
	G=100, n=10	0.207	0.064	0.025	36.955	0.212	0.005	0.843	0.186	0.017	0.820	0.448	0.036
	G=100, n=30	0.124	0.025	0.013	37.917	0.265	0.004	0.560	0.225	0.009	0.953	0.427	0.023
ICC=0.05	G=30, n=10	0.178	0.039	0.020	39.737	0.224	0.006	0.724	0.161	0.018	1.093	0.431	0.033
	G=30, n=30	0.079	0.014	0.008	31.087	0.254	0.004	0.572	0.202	0.008	1.129	0.417	0.017
	G=100, n=10	0.108	0.013	0.008	40.385	0.275	0.003	0.633	0.216	0.008	7.892	0.414	0.016
	G=100, n=30	0.032	0.004	0.003	26.548	0.226	0.002	0.348	0.186	0.004	0.886	0.367	0.010
ICC=0.40	G=30, n=10	0.172	0.042	0.021	37.687	0.232	0.006	0.749	0.163	0.019	0.300	0.173	0.016
	G=30, n=30	0.079	0.013	0.009	30.460	0.267	0.004	0.576	0.187	0.009	0.331	0.172	0.009
	G=100, n=10	0.091	0.013	0.008	34.827	0.240	0.004	0.587	0.189	0.008	0.481	0.155	0.007
	G=100, n=30	0.029	0.004	0.003	24.729	0.234	0.002	0.327	0.180	0.004	0.424	0.137	0.005

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 7 **Error cuadrático medio** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 5)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.411	0.195	0.055	36.574	0.227	0.007	1.571	0.243	0.048	0.798	0.436	0.080
	G=30, n=30	0.338	0.080	0.033	37.583	0.249	0.006	1.400	0.306	0.023	0.866	0.453	0.045
	G=100, n=10	0.326	0.080	0.030	43.734	0.255	0.005	1.593	0.380	0.020	0.836	0.418	0.041
	G=100, n=30	0.166	0.025	0.013	42.961	0.276	0.004	1.186	0.377	0.009	1.036	0.449	0.024
ICC=0.05	G=30, n=10	0.286	0.044	0.021	48.632	0.230	0.006	1.405	0.285	0.020	1.167	0.438	0.036
	G=30, n=30	0.130	0.015	0.009	41.984	0.278	0.004	1.135	0.347	0.008	1.708	0.432	0.019
	G=100, n=10	0.150	0.014	0.008	50.769	0.295	0.003	1.282	0.387	0.008	0.927	0.432	0.016
	G=100, n=30	0.042	0.004	0.003	30.762	0.267	0.002	0.858	0.302	0.004	1.196	0.361	0.009
ICC=0.40	G=30, n=10	0.265	0.044	0.023	44.695	0.264	0.006	1.434	0.284	0.021	0.340	0.185	0.017
	G=30, n=30	0.125	0.014	0.009	44.127	0.280	0.004	1.120	0.321	0.009	0.331	0.157	0.009
	G=100, n=10	0.124	0.013	0.008	44.168	0.273	0.004	1.243	0.332	0.008	0.387	0.163	0.007
	G=100, n=30	0.044	0.004	0.003	32.072	0.259	0.002	0.717	0.361	0.004	0.594	0.186	0.004

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 8 **Error cuadrático medio** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 6)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	7.351	2.060	0.114	97.583	0.352	0.009	7.012	1.834	0.101	0.905	0.518	0.116
	G=30, n=30	2.620	0.230	0.047	89.692	0.371	0.008	6.597	1.146	0.032	0.923	0.467	0.062
	G=100, n=10	2.732	0.151	0.041	98.985	0.398	0.007	6.496	1.185	0.030	1.066	0.461	0.062
	G=100, n=30	0.794	0.034	0.016	84.582	0.397	0.004	5.822	1.002	0.011	1.912	0.428	0.026
ICC=0.05	G=30, n=10	1.450	0.066	0.029	94.070	0.391	0.007	5.680	0.971	0.028	1.214	0.452	0.053
	G=30, n=30	0.316	0.016	0.011	65.673	0.358	0.005	4.493	0.908	0.011	0.944	0.431	0.025
	G=100, n=10	0.424	0.016	0.010	89.921	0.411	0.005	5.878	1.053	0.010	1.156	0.412	0.024
	G=100, n=30	0.056	0.004	0.003	37.635	0.360	0.003	3.631	0.783	0.004	0.742	0.451	0.011
ICC=0.40	G=30, n=10	1.435	0.061	0.028	95.664	0.340	0.007	5.840	0.951	0.026	0.405	0.182	0.020
	G=30, n=30	0.393	0.018	0.010	79.984	0.371	0.004	5.449	0.854	0.011	0.364	0.163	0.010
	G=100, n=10	0.348	0.016	0.009	83.738	0.365	0.004	5.473	0.939	0.009	0.335	0.176	0.008
	G=100, n=30	0.056	0.004	0.003	32.554	0.342	0.002	3.890	0.821	0.004	0.427	0.181	0.005

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

5.1.1.2.2 Nivel superior (dentro de grupos)

Los resultados encontrados en este nivel son similares a los presentados para el nivel dentro de grupos. A lo largo de todas las condiciones de estimación que fueron simuladas, las estimaciones Bayesianas con distribuciones a priori débilmente informativas e informativas en las cargas factoriales cruzadas mostraron valores pequeños del RMSE consistentemente. Para la estimación Bayesiana con distribución a priori difusa, el RMSE estuvo negativamente asociado al número de grupos, el tamaño de los grupos y el ICC.

La carga factorial cruzada fue la única que presentó altos valores de RMSE para el escenario de mayor información suministrada al modelo (escenario 1) mientras que entre menor información se ingresaba al mismo en distintos parámetros, se presentaban altos valores también para el parámetro estructural. Sólo en el caso del escenario 6 (el de menor información) la carga factorial directa también mostró altos valores de RMSE.

Tabla 9 **Error cuadrático medio** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 1)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.082	0.041	0.030	19.987	0.266	0.006	0.076	0.040	0.026	0.032	0.028	0.026
	G=30, n=30	0.066	0.035	0.031	18.218	0.249	0.005	0.063	0.030	0.027	0.013	0.013	0.011
	G=100, n=10	0.063	0.045	0.032	18.129	0.370	0.004	0.067	0.040	0.026	0.010	0.009	0.008
	G=100, n=30	0.050	0.038	0.029	22.406	0.259	0.004	0.064	0.034	0.034	0.004	0.004	0.004
ICC=0.05	G=30, n=10	0.073	0.038	0.034	17.392	0.312	0.005	0.074	0.031	0.026	0.024	0.022	0.018
	G=30, n=30	0.061	0.034	0.034	16.750	0.301	0.005	0.068	0.032	0.025	0.013	0.011	0.009
	G=100, n=10	0.059	0.037	0.034	15.441	0.374	0.004	0.068	0.038	0.028	0.007	0.006	0.006
	G=100, n=30	0.048	0.035	0.032	21.838	0.239	0.004	0.060	0.029	0.032	0.004	0.003	0.003
ICC=0.40	G=30, n=10	0.077	0.046	0.042	53.849	0.312	0.006	0.080	0.038	0.039	0.065	0.051	0.048
	G=30, n=30	0.066	0.039	0.038	54.035	0.289	0.006	0.078	0.035	0.038	0.058	0.048	0.043
	G=100, n=10	0.054	0.041	0.028	53.362	0.222	0.004	0.098	0.032	0.033	0.046	0.015	0.014
	G=100, n=30	0.046	0.037	0.028	47.669	0.203	0.005	0.091	0.028	0.030	0.045	0.014	0.012

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 10 **Error cuadrático medio** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 2)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.203	0.135	0.115	4.863	0.222	0.007	0.269	0.144	0.114	0.034	0.030	0.028
	G=30, n=30	0.148	0.107	0.097	3.688	0.214	0.006	0.265	0.123	0.099	0.015	0.014	0.012
	G=100, n=10	0.183	0.136	0.127	3.675	0.246	0.005	0.346	0.176	0.116	0.010	0.009	0.008
	G=100, n=30	0.137	0.115	0.079	3.734	0.229	0.005	0.489	0.150	0.091	0.004	0.004	0.004
ICC=0.05	G=30, n=10	0.179	0.128	0.132	4.376	0.217	0.006	0.253	0.134	0.110	0.025	0.022	0.019
	G=30, n=30	0.152	0.133	0.133	3.317	0.241	0.005	0.311	0.140	0.117	0.012	0.011	0.010
	G=100, n=10	0.182	0.146	0.147	2.918	0.274	0.005	0.372	0.189	0.135	0.007	0.007	0.006
	G=100, n=30	0.165	0.146	0.102	3.148	0.216	0.004	0.473	0.160	0.103	0.004	0.004	0.003
ICC=0.40	G=30, n=10	0.239	0.174	0.148	22.025	0.307	0.006	0.407	0.134	0.138	0.062	0.052	0.047
	G=30, n=30	0.197	0.172	0.131	22.100	0.280	0.006	0.432	0.135	0.119	0.055	0.046	0.044
	G=100, n=10	0.208	0.190	0.091	19.414	0.292	0.005	0.742	0.164	0.086	0.038	0.016	0.014
	G=100, n=30	0.178	0.184	0.075	18.160	0.253	0.004	0.762	0.141	0.072	0.038	0.014	0.012

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 11 **Error cuadrático medio** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 3)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.319	0.220	0.209	4.816	0.220	0.008	0.490	0.268	0.200	0.035	0.031	0.027
	G=30, n=30	0.256	0.175	0.157	4.295	0.222	0.006	0.523	0.208	0.173	0.015	0.014	0.012
	G=100, n=10	0.328	0.248	0.211	4.059	0.279	0.006	0.633	0.316	0.215	0.010	0.009	0.008
	G=100, n=30	0.240	0.171	0.111	4.207	0.251	0.004	0.820	0.232	0.129	0.004	0.004	0.004
ICC=0.05	G=30, n=10	0.294	0.228	0.258	4.524	0.226	0.006	0.506	0.277	0.224	0.025	0.023	0.020
	G=30, n=30	0.269	0.236	0.236	3.870	0.243	0.006	0.560	0.247	0.206	0.013	0.012	0.010
	G=100, n=10	0.342	0.287	0.284	3.272	0.288	0.005	0.666	0.367	0.242	0.007	0.007	0.006
	G=100, n=30	0.252	0.243	0.158	3.542	0.265	0.004	0.819	0.262	0.145	0.004	0.003	0.003
ICC=0.40	G=30, n=10	0.464	0.327	0.242	21.767	0.350	0.007	0.697	0.218	0.202	0.066	0.054	0.048
	G=30, n=30	0.385	0.330	0.218	21.189	0.353	0.006	0.745	0.225	0.181	0.056	0.051	0.045
	G=100, n=10	0.373	0.367	0.148	19.767	0.334	0.005	1.212	0.248	0.126	0.039	0.016	0.013
	G=100, n=30	0.356	0.340	0.113	19.384	0.356	0.004	1.278	0.233	0.104	0.039	0.015	0.013

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 12 **Error cuadrático medio** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 4)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.271	0.127	0.111	7.279	0.178	0.007	0.345	0.153	0.113	0.036	0.032	0.028
	G=30, n=30	0.204	0.102	0.094	8.283	0.189	0.006	0.270	0.118	0.094	0.016	0.015	0.012
	G=100, n=10	0.225	0.141	0.114	9.947	0.242	0.005	0.371	0.203	0.110	0.010	0.009	0.008
	G=100, n=30	0.166	0.113	0.073	10.923	0.227	0.004	0.378	0.186	0.086	0.004	0.004	0.004
ICC=0.05	G=30, n=10	0.238	0.117	0.110	7.819	0.181	0.006	0.327	0.158	0.103	0.026	0.023	0.020
	G=30, n=30	0.197	0.112	0.121	7.648	0.199	0.005	0.295	0.153	0.104	0.013	0.012	0.011
	G=100, n=10	0.204	0.144	0.127	9.068	0.237	0.005	0.338	0.188	0.117	0.008	0.006	0.006
	G=100, n=30	0.151	0.122	0.093	10.277	0.199	0.004	0.394	0.177	0.098	0.004	0.003	0.003
ICC=0.40	G=30, n=10	0.193	0.123	0.110	11.769	0.202	0.007	0.360	0.151	0.108	0.065	0.053	0.047
	G=30, n=30	0.152	0.116	0.102	9.303	0.202	0.006	0.356	0.147	0.093	0.060	0.048	0.044
	G=100, n=10	0.158	0.111	0.067	14.390	0.188	0.005	0.546	0.165	0.064	0.034	0.015	0.014
	G=100, n=30	0.151	0.110	0.059	13.169	0.159	0.004	0.532	0.139	0.058	0.032	0.015	0.013

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 13 **Error cuadrático medio** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 5)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.418	0.206	0.184	7.185	0.198	0.008	0.576	0.265	0.187	0.034	0.030	0.028
	G=30, n=30	0.316	0.173	0.144	9.129	0.190	0.006	0.528	0.233	0.153	0.015	0.014	0.012
	G=100, n=10	0.381	0.239	0.183	10.926	0.297	0.005	0.605	0.383	0.178	0.009	0.010	0.008
	G=100, n=30	0.304	0.167	0.100	12.460	0.260	0.004	0.673	0.320	0.120	0.004	0.004	0.004
ICC=0.05	G=30, n=10	0.380	0.186	0.190	8.444	0.192	0.006	0.500	0.240	0.186	0.025	0.023	0.019
	G=30, n=30	0.319	0.206	0.185	8.034	0.207	0.006	0.481	0.290	0.170	0.013	0.012	0.010
	G=100, n=10	0.368	0.252	0.210	9.976	0.258	0.005	0.548	0.377	0.200	0.007	0.007	0.006
	G=100, n=30	0.282	0.209	0.127	10.630	0.219	0.004	0.713	0.312	0.138	0.004	0.004	0.003
ICC=0.40	G=30, n=10	0.332	0.207	0.149	11.979	0.207	0.007	0.612	0.256	0.152	0.070	0.055	0.048
	G=30, n=30	0.309	0.188	0.135	10.857	0.201	0.007	0.648	0.234	0.123	0.059	0.050	0.043
	G=100, n=10	0.278	0.190	0.093	15.231	0.209	0.005	0.913	0.248	0.086	0.034	0.017	0.014
	G=100, n=30	0.237	0.168	0.076	13.288	0.197	0.004	0.955	0.219	0.070	0.033	0.014	0.013

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

Tabla 14 **Error cuadrático medio** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 6)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	10.383	2.269	2.831	15.898	0.293	0.010	3.415	1.694	2.191	0.034	0.033	0.034
	G=30, n=30	9.671	1.913	2.481	26.531	0.289	0.009	3.054	1.622	2.048	0.015	0.015	0.013
	G=100, n=10	16.148	3.266	4.265	31.415	0.377	0.008	3.504	2.268	3.371	0.009	0.010	0.009
	G=100, n=30	10.591	2.088	1.303	38.813	0.349	0.005	3.247	2.007	1.462	0.004	0.004	0.004
ICC=0.05	G=30, n=10	10.200	2.137	3.680	17.448	0.289	0.009	2.947	1.623	2.391	0.024	0.023	0.023
	G=30, n=30	11.261	2.544	3.537	25.402	0.293	0.009	2.609	1.450	2.123	0.012	0.013	0.012
	G=100, n=10	15.415	3.193	6.126	27.403	0.349	0.008	3.044	2.040	3.493	0.007	0.007	0.007
	G=100, n=30	11.251	3.198	2.648	40.938	0.316	0.006	2.845	1.985	1.668	0.004	0.004	0.003
ICC=0.40	G=30, n=10	10.273	2.326	2.747	38.247	0.324	0.008	2.246	1.498	2.048	0.076	0.064	0.055
	G=30, n=30	9.213	1.944	1.923	41.489	0.291	0.008	2.155	1.377	1.404	0.060	0.055	0.047
	G=100, n=10	6.041	2.468	0.654	42.669	0.269	0.006	2.821	1.716	0.560	0.044	0.022	0.015
	G=100, n=30	5.160	2.213	0.220	49.227	0.255	0.005	2.650	1.411	0.215	0.044	0.020	0.013

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los errores cuadráticos medios superiores a 0.5

5.1.2 Sesgos en las estimaciones de los parámetros

Las estimaciones frecuentistas usualmente producen bajos niveles de recuperación del parámetro en valores bajos del ICC, y mejoran cuando este valor se incrementa hasta cierto umbral, ya que en valores altos de ICC también muestra malos ajustes (Depaoli & Clifton, 2015). Para la estimación Bayesiana, la cobertura estuvo inversamente relacionada a la fuerza del ICC (se obtuvieron mayores niveles de recuperación de los parámetros en los niveles inferiores del ICC y a menores tamaños de muestra).

Nivel inferior (dentro de grupos)

En el caso de las cargas factoriales directas, los niveles de recuperación fueron consistentemente altos (superiores a 0.95) en la mayoría de escenarios simulados. Pero para las cargas factoriales cruzadas y el parámetro estructural, únicamente los casos con distribución a priori informativa en la carga factorial cruzada muestran niveles aceptables de recuperación del parámetro en todos los escenarios. El parámetro de varianza residual y factorial no logró obtener niveles aceptables de recuperación del parámetro para los escenarios simulados.

Tabla 15 Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 1)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.978	0.983	0.981	0.136	0.908	0.999	0.974	1.000	0.994	0.148	0.902	0.956
	G=30, n=30	0.966	0.969	0.986	0.172	0.739	0.998	0.950	0.982	0.991	0.196	0.740	0.948
	G=100, n=10	0.963	0.969	0.982	0.108	0.889	0.998	0.959	0.999	0.995	0.135	0.890	0.958
	G=100, n=30	0.962	0.958	0.963	0.178	0.640	0.999	0.930	0.972	0.989	0.200	0.651	0.944
ICC=0.05	G=30, n=10	0.962	0.961	0.972	0.153	0.782	0.998	0.954	0.997	0.986	0.172	0.789	0.953
	G=30, n=30	0.959	0.956	0.948	0.216	0.540	0.995	0.891	0.927	0.987	0.253	0.553	0.944
	G=100, n=10	0.958	0.942	0.944	0.166	0.638	0.998	0.935	0.963	0.991	0.186	0.651	0.945
	G=100, n=30	0.951	0.950	0.943	0.262	0.429	0.978	0.850	0.825	0.990	0.269	0.430	0.933
ICC=0.40	G=30, n=10	0.966	0.964	0.965	0.156	0.726	0.996	0.958	0.982	0.985	0.183	0.739	0.950
	G=30, n=30	0.956	0.949	0.946	0.215	0.535	0.993	0.894	0.915	0.985	0.253	0.542	0.936
	G=100, n=10	0.967	0.946	0.933	0.209	0.577	0.975	0.883	0.947	0.981	0.244	0.577	0.938
	G=100, n=30	0.953	0.949	0.942	0.294	0.431	0.954	0.800	0.822	0.984	0.310	0.444	0.926

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura mayores a 0.95

Tabla 16 **Porcentaje de recuperación** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 2)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.975	0.965	0.971	0.048	0.789	0.999	0.658	0.988	0.977	0.100	0.762	0.948
	G=30, n=30	0.964	0.948	0.962	0.073	0.635	0.992	0.691	0.923	0.971	0.121	0.621	0.944
	G=100, n=10	0.969	0.946	0.967	0.024	0.733	0.995	0.659	0.946	0.976	0.068	0.744	0.949
	G=100, n=30	0.960	0.949	0.942	0.098	0.524	0.984	0.750	0.876	0.976	0.139	0.554	0.929
ICC=0.05	G=30, n=10	0.960	0.954	0.957	0.038	0.656	0.999	0.709	0.929	0.980	0.058	0.690	0.950
	G=30, n=30	0.961	0.949	0.941	0.141	0.485	0.988	0.691	0.783	0.982	0.194	0.501	0.931
	G=100, n=10	0.960	0.936	0.935	0.085	0.534	0.995	0.709	0.881	0.989	0.125	0.557	0.945
	G=100, n=30	0.955	0.953	0.943	0.201	0.407	0.974	0.700	0.684	0.988	0.247	0.422	0.927
ICC=0.40	G=30, n=10	0.967	0.952	0.954	0.051	0.664	0.993	0.699	0.925	0.974	0.099	0.705	0.949
	G=30, n=30	0.955	0.950	0.942	0.139	0.506	0.983	0.697	0.801	0.978	0.188	0.541	0.941
	G=100, n=10	0.959	0.934	0.923	0.093	0.545	0.985	0.710	0.856	0.974	0.134	0.560	0.937
	G=100, n=30	0.941	0.956	0.948	0.206	0.423	0.953	0.689	0.702	0.980	0.264	0.432	0.925

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

Tabla 17 **Porcentaje de recuperación** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 3)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.984	0.955	0.972	0.031	0.765	0.998	0.492	0.978	0.981	0.065	0.767	0.949
	G=30, n=30	0.969	0.931	0.964	0.049	0.572	0.983	0.572	0.915	0.967	0.067	0.604	0.938
	G=100, n=10	0.973	0.938	0.956	0.016	0.677	0.992	0.507	0.946	0.981	0.037	0.713	0.949
	G=100, n=30	0.956	0.951	0.941	0.095	0.521	0.983	0.696	0.833	0.970	0.125	0.561	0.929
ICC=0.05	G=30, n=10	0.968	0.944	0.958	0.025	0.631	0.994	0.536	0.919	0.977	0.048	0.679	0.946
	G=30, n=30	0.956	0.945	0.936	0.116	0.470	0.989	0.609	0.786	0.986	0.157	0.525	0.941
	G=100, n=10	0.955	0.935	0.931	0.044	0.556	0.998	0.612	0.852	0.984	0.076	0.597	0.945
	G=100, n=30	0.957	0.955	0.944	0.207	0.438	0.981	0.644	0.680	0.991	0.263	0.463	0.930
ICC=0.40	G=30, n=10	0.958	0.952	0.952	0.022	0.619	0.993	0.542	0.918	0.967	0.053	0.673	0.946
	G=30, n=30	0.961	0.945	0.949	0.084	0.486	0.988	0.606	0.779	0.980	0.121	0.534	0.943
	G=100, n=10	0.952	0.937	0.933	0.083	0.577	0.993	0.601	0.832	0.978	0.125	0.618	0.937
	G=100, n=30	0.941	0.958	0.948	0.213	0.426	0.956	0.672	0.689	0.983	0.254	0.453	0.931

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

Tabla 18 **Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 4)**

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.977	0.973	0.974	0.045	0.802	0.997	0.658	0.978	0.982	0.099	0.780	0.947
	G=30, n=30	0.963	0.947	0.962	0.072	0.619	0.994	0.674	0.925	0.964	0.126	0.624	0.933
	G=100, n=10	0.972	0.945	0.966	0.032	0.745	0.996	0.669	0.971	0.986	0.075	0.741	0.952
	G=100, n=30	0.964	0.951	0.949	0.092	0.580	0.990	0.774	0.906	0.977	0.117	0.598	0.935
ICC=0.05	G=30, n=10	0.963	0.948	0.956	0.055	0.661	0.996	0.683	0.937	0.976	0.112	0.702	0.944
	G=30, n=30	0.960	0.943	0.942	0.170	0.470	0.989	0.686	0.807	0.980	0.204	0.507	0.938
	G=100, n=10	0.963	0.944	0.930	0.086	0.572	0.993	0.711	0.872	0.985	0.127	0.599	0.942
	G=100, n=30	0.947	0.952	0.948	0.203	0.451	0.970	0.732	0.693	0.990	0.237	0.467	0.928
ICC=0.40	G=30, n=10	0.972	0.942	0.947	0.047	0.595	0.997	0.679	0.898	0.968	0.101	0.645	0.944
	G=30, n=30	0.952	0.944	0.945	0.168	0.457	0.989	0.690	0.784	0.977	0.218	0.490	0.937
	G=100, n=10	0.959	0.935	0.933	0.091	0.566	0.989	0.715	0.847	0.979	0.126	0.589	0.939
	G=100, n=30	0.958	0.951	0.943	0.206	0.447	0.948	0.733	0.690	0.984	0.253	0.454	0.937

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

Tabla 19 **Porcentaje de recuperación de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 5)**

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.976	0.961	0.970	0.041	0.747	0.999	0.489	0.974	0.967	0.069	0.729	0.952
	G=30, n=30	0.973	0.939	0.959	0.058	0.584	0.982	0.550	0.910	0.960	0.096	0.631	0.929
	G=100, n=10	0.977	0.939	0.964	0.025	0.750	0.992	0.527	0.961	0.982	0.043	0.782	0.944
	G=100, n=30	0.950	0.951	0.943	0.070	0.553	0.993	0.660	0.877	0.977	0.106	0.600	0.932
ICC=0.05	G=30, n=10	0.967	0.941	0.955	0.027	0.649	0.996	0.558	0.921	0.972	0.079	0.702	0.949
	G=30, n=30	0.950	0.935	0.941	0.122	0.499	0.991	0.586	0.779	0.980	0.180	0.534	0.937
	G=100, n=10	0.957	0.939	0.939	0.064	0.547	0.998	0.599	0.874	0.983	0.108	0.596	0.944
	G=100, n=30	0.955	0.954	0.943	0.168	0.424	0.979	0.642	0.689	0.987	0.237	0.444	0.930
ICC=0.40	G=30, n=10	0.970	0.946	0.944	0.032	0.608	0.992	0.531	0.887	0.976	0.064	0.650	0.943
	G=30, n=30	0.952	0.948	0.939	0.098	0.468	0.989	0.623	0.767	0.978	0.132	0.526	0.934
	G=100, n=10	0.959	0.937	0.938	0.074	0.518	0.987	0.611	0.848	0.978	0.098	0.571	0.937
	G=100, n=30	0.945	0.953	0.943	0.188	0.390	0.958	0.705	0.661	0.988	0.236	0.432	0.929

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

Tabla 20 **Porcentaje de recuperación** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos, escenario 6)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	0.958	0.879	0.961	0.011	0.564	0.994	0.246	0.949	0.957	0.019	0.592	0.948
	G=30, n=30	0.948	0.917	0.963	0.029	0.488	0.986	0.210	0.846	0.964	0.057	0.595	0.939
	G=100, n=10	0.931	0.931	0.958	0.011	0.639	0.988	0.149	0.914	0.972	0.028	0.736	0.947
	G=100, n=30	0.938	0.943	0.948	0.057	0.462	0.993	0.242	0.796	0.976	0.090	0.582	0.933
ICC=0.05	G=30, n=10	0.945	0.941	0.956	0.031	0.552	0.997	0.204	0.871	0.971	0.051	0.687	0.949
	G=30, n=30	0.955	0.939	0.952	0.109	0.442	0.988	0.278	0.662	0.981	0.158	0.567	0.943
	G=100, n=10	0.944	0.938	0.940	0.065	0.514	0.996	0.229	0.793	0.983	0.103	0.637	0.942
	G=100, n=30	0.944	0.952	0.945	0.217	0.381	0.988	0.430	0.588	0.995	0.282	0.465	0.927
ICC=0.40	G=30, n=10	0.949	0.939	0.948	0.028	0.544	0.994	0.173	0.873	0.972	0.049	0.674	0.941
	G=30, n=30	0.953	0.941	0.939	0.077	0.442	0.992	0.237	0.711	0.979	0.132	0.545	0.935
	G=100, n=10	0.928	0.939	0.933	0.049	0.534	0.988	0.195	0.821	0.979	0.080	0.646	0.940
	G=100, n=30	0.941	0.957	0.943	0.159	0.376	0.962	0.374	0.626	0.981	0.244	0.476	0.925

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

Nivel superior (dentro de grupos)

Para este nivel se puede apreciar que la cobertura fue buena para todos los niveles de distribución a priori débilmente informativa a valores bajos del ICC (Tablas 19 a 24). Sin embargo, esta cobertura disminuye conforme el ICC se incrementa, siendo más pronunciado este efecto para las distribuciones a priori débilmente informativas en donde las varianzas del nivel *entre grupos* fueron especificadas como $\Gamma^{-1}(0.001,0.001)$ comparadas con las distribuciones a priori débilmente informativas en donde la dicha varianza fue especificada como $\Gamma^{-1}(-1,0)$.

Tabla 21 **Porcentaje de recuperación** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 1)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	1.000	1.000	1.000	0.975	0.998	1.000	1.000	1.000	1.000	0.001	0.906	0.888
	G=30, n=30	1.000	1.000	0.999	0.853	0.990	1.000	1.000	1.000	1.000	0.002	0.723	0.902
	G=100, n=10	1.000	1.000	0.999	0.902	0.990	1.000	1.000	1.000	1.000	0.000	0.664	0.901
	G=100, n=30	0.999	1.000	0.991	0.357	0.939	1.000	0.997	1.000	0.997	0.004	0.742	0.920
ICC=0.05	G=30, n=10	1.000	1.000	0.995	0.955	0.992	1.000	1.000	1.000	1.000	0.002	0.905	0.892
	G=30, n=30	1.000	1.000	0.999	0.897	0.984	1.000	1.000	1.000	0.999	0.000	0.914	0.920
	G=100, n=10	1.000	1.000	0.989	0.910	0.983	1.000	1.000	1.000	1.000	0.004	0.716	0.913
	G=100, n=30	0.999	0.999	0.988	0.541	0.948	1.000	0.999	1.000	0.997	0.005	0.736	0.935
ICC=0.40	G=30, n=10	0.999	1.000	0.987	0.514	0.960	1.000	0.999	1.000	0.994	0.001	0.862	0.929
	G=30, n=30	0.999	1.000	0.987	0.367	0.969	1.000	1.000	1.000	0.998	0.000	0.904	0.933
	G=100, n=10	0.991	0.996	0.983	0.072	0.904	1.000	0.968	0.999	0.981	0.008	0.810	0.940
	G=100, n=30	0.995	0.992	0.988	0.014	0.920	1.000	0.958	1.000	0.989	0.013	0.862	0.940

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura mayores a 0.95

Tabla 22 **Porcentaje de recuperación** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 2)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	1.000	1.000	0.999	0.977	1.000	1.000	1.000	1.000	1.000	0.538	0.934	0.928
	G=30, n=30	1.000	1.000	0.997	0.900	0.998	1.000	1.000	1.000	0.999	0.335	0.933	0.929
	G=100, n=10	1.000	1.000	0.995	0.916	0.994	1.000	0.998	1.000	0.998	0.291	0.931	0.919
	G=100, n=30	0.999	0.997	0.982	0.662	0.966	1.000	0.978	1.000	0.986	0.218	0.865	0.928
ICC=0.05	G=30, n=10	1.000	1.000	0.994	0.977	0.999	1.000	1.000	1.000	0.999	0.754	0.928	0.910
	G=30, n=30	1.000	1.000	0.989	0.946	0.991	1.000	0.999	1.000	0.997	0.684	0.929	0.925
	G=100, n=10	1.000	0.999	0.983	0.945	0.987	1.000	0.998	1.000	0.995	0.475	0.936	0.922
	G=100, n=30	0.996	0.998	0.972	0.791	0.969	1.000	0.970	0.999	0.987	0.403	0.933	0.931
ICC=0.40	G=30, n=10	0.999	0.998	0.973	0.525	0.978	1.000	0.975	1.000	0.986	0.026	0.915	0.931
	G=30, n=30	1.000	0.996	0.967	0.379	0.975	1.000	0.974	1.000	0.993	0.018	0.933	0.933
	G=100, n=10	0.998	0.984	0.959	0.109	0.954	1.000	0.817	1.000	0.978	0.017	0.869	0.923
	G=100, n=30	0.994	0.985	0.969	0.089	0.948	1.000	0.789	1.000	0.972	0.019	0.898	0.939

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

Tabla 23 **Porcentaje de recuperación** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 3)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	1.000	1.000	0.998	0.983	1.000	1.000	1.000	1.000	1.000	0.521	0.933	0.929
	G=30, n=30	1.000	1.000	0.994	0.922	0.999	1.000	0.999	1.000	0.998	0.289	0.932	0.921
	G=100, n=10	1.000	0.999	0.995	0.933	0.992	1.000	0.999	1.000	0.996	0.248	0.935	0.926
	G=100, n=30	0.996	0.997	0.975	0.645	0.966	1.000	0.960	0.999	0.986	0.189	0.845	0.912
ICC=0.05	G=30, n=10	1.000	0.999	0.993	0.978	0.997	1.000	0.999	1.000	0.999	0.736	0.921	0.911
	G=30, n=30	1.000	1.000	0.986	0.946	0.995	1.000	0.998	0.999	0.994	0.601	0.930	0.929
	G=100, n=10	1.000	0.999	0.971	0.943	0.994	1.000	0.998	1.000	0.987	0.442	0.939	0.932
	G=100, n=30	0.997	0.997	0.966	0.776	0.974	1.000	0.960	0.999	0.976	0.348	0.924	0.932
ICC=0.40	G=30, n=10	0.998	0.995	0.958	0.516	0.974	1.000	0.945	1.000	0.983	0.012	0.861	0.916
	G=30, n=30	1.000	0.995	0.969	0.403	0.968	1.000	0.931	1.000	0.989	0.011	0.887	0.934
	G=100, n=10	0.999	0.972	0.942	0.071	0.931	1.000	0.684	0.999	0.957	0.011	0.813	0.896
	G=100, n=30	0.998	0.982	0.963	0.044	0.940	1.000	0.651	0.999	0.960	0.012	0.837	0.912

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

Tabla 24 **Porcentaje de recuperación** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 4)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	1.000	1.000	0.999	0.994	0.999	1.000	1.000	1.000	0.998	0.168	0.950	0.928
	G=30, n=30	1.000	1.000	0.997	0.963	0.996	1.000	1.000	1.000	0.999	0.112	0.943	0.939
	G=100, n=10	1.000	1.000	0.994	0.964	0.993	1.000	1.000	1.000	0.998	0.086	0.937	0.925
	G=100, n=30	1.000	0.999	0.984	0.657	0.971	1.000	0.992	1.000	0.989	0.074	0.904	0.929
ICC=0.05	G=30, n=10	1.000	1.000	0.993	0.987	1.000	1.000	1.000	1.000	0.998	0.173	0.940	0.930
	G=30, n=30	1.000	1.000	0.991	0.971	0.996	1.000	1.000	1.000	0.998	0.154	0.941	0.942
	G=100, n=10	1.000	0.999	0.984	0.961	0.995	1.000	1.000	1.000	0.993	0.092	0.937	0.935
	G=100, n=30	0.999	0.997	0.975	0.747	0.974	1.000	0.994	0.999	0.987	0.076	0.928	0.938
ICC=0.40	G=30, n=10	0.999	1.000	0.979	0.763	0.995	1.000	0.998	1.000	0.989	0.184	0.951	0.952
	G=30, n=30	1.000	1.000	0.987	0.776	0.991	1.000	0.997	1.000	0.992	0.208	0.952	0.946
	G=100, n=10	1.000	0.994	0.969	0.314	0.977	1.000	0.926	1.000	0.980	0.109	0.946	0.942
	G=100, n=30	0.997	0.995	0.976	0.270	0.974	1.000	0.897	1.000	0.974	0.121	0.945	0.940

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

Tabla 25 **Porcentaje de recuperación** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos, escenario 5)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	1.000	1.000	0.998	0.996	1.000	1.000	1.000	1.000	0.999	0.139	0.952	0.928
	G=30, n=30	1.000	1.000	0.998	0.964	0.998	1.000	1.000	1.000	0.998	0.094	0.939	0.941
	G=100, n=10	1.000	0.999	0.988	0.978	0.996	1.000	1.000	1.000	0.998	0.071	0.914	0.935
	G=100, n=30	0.997	0.998	0.979	0.687	0.970	1.000	0.986	1.000	0.989	0.064	0.855	0.932
ICC=0.05	G=30, n=10	1.000	1.000	0.995	0.987	0.997	1.000	1.000	1.000	0.995	0.132	0.939	0.925
	G=30, n=30	1.000	1.000	0.987	0.967	0.998	1.000	1.000	1.000	0.994	0.120	0.939	0.934
	G=100, n=10	1.000	0.999	0.973	0.976	0.991	1.000	1.000	1.000	0.995	0.067	0.919	0.929
	G=100, n=30	1.000	0.997	0.978	0.773	0.969	1.000	0.981	1.000	0.978	0.079	0.898	0.936
ICC=0.40	G=30, n=10	1.000	0.999	0.980	0.820	0.989	1.000	0.993	1.000	0.987	0.186	0.954	0.942
	G=30, n=30	1.000	0.999	0.980	0.804	0.993	1.000	0.994	1.000	0.993	0.163	0.951	0.947
	G=100, n=10	0.998	0.989	0.961	0.362	0.958	1.000	0.877	1.000	0.979	0.092	0.932	0.946
	G=100, n=30	0.994	0.991	0.972	0.284	0.954	1.000	0.848	0.999	0.971	0.080	0.932	0.942

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

Tabla 26 **Porcentaje de recuperación** de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro entre grupos, escenario 6)

ICC	N	Carga factorial directa			Carga factorial cruzada			Parámetro estructural			Varianza residual		
		Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte	Difusa	Débil	Fuerte
ICC=0.02	G=30, n=10	1.000	1.000	0.997	0.998	0.998	1.000	1.000	1.000	1.000	0.007	0.374	0.453
	G=30, n=30	1.000	0.999	0.995	0.983	0.969	1.000	0.999	1.000	0.997	0.003	0.208	0.372
	G=100, n=10	0.999	0.997	0.988	0.988	0.978	1.000	0.999	1.000	0.994	0.002	0.129	0.268
	G=100, n=30	0.991	0.986	0.959	0.712	0.679	1.000	0.935	0.997	0.950	0.002	0.173	0.618
ICC=0.05	G=30, n=10	1.000	0.998	0.992	0.998	0.995	1.000	1.000	1.000	0.999	0.003	0.333	0.424
	G=30, n=30	0.999	0.995	0.983	0.977	0.971	1.000	1.000	1.000	0.994	0.004	0.264	0.427
	G=100, n=10	0.999	0.987	0.966	0.988	0.980	1.000	0.997	1.000	0.986	0.002	0.174	0.275
	G=100, n=30	0.991	0.978	0.954	0.797	0.791	1.000	0.940	1.000	0.959	0.004	0.243	0.547
ICC=0.40	G=30, n=10	0.995	0.982	0.955	0.830	0.821	1.000	0.985	0.998	0.951	0.005	0.369	0.624
	G=30, n=30	0.994	0.982	0.952	0.809	0.770	1.000	0.976	1.000	0.970	0.001	0.417	0.742
	G=100, n=10	0.989	0.964	0.911	0.272	0.599	0.998	0.722	0.996	0.918	0.009	0.432	0.821
	G=100, n=30	0.982	0.948	0.937	0.134	0.566	0.999	0.598	1.000	0.931	0.002	0.449	0.893

ICC: Coeficiente de correlación intraclase; N= Tamaño de muestra definido por el número de grupos (G) y el tamaño de grupo (n).

Las columnas Difusa, Débil y Fuerte muestran el nivel de información en la varianza de la distribución a priori de las cargas factoriales cruzadas.

En negrita se resaltan los porcentajes de cobertura menores a 0.95

5.1.3 Bondad de ajuste de los modelos

Como fue explicado en la sección 4.1.4, para evaluar el ajuste global del modelo se recurrirá al cómputo de la probabilidad asociada a la prueba cociente de verosimilitudes del modelo nulo respecto del modelo irrestricto haciendo uso del *P-value predictivo a posteriori (PPP)*. Esta probabilidad no se comporta como un p-value para una prueba chi-cuadrado de ajuste de modelos, sino que es más similar a un índice de ajuste de un SEM (como el RMSEA) que a una prueba chi-cuadrado.

En la actualidad, no existe teoría aún sobre cuáles podrían ser los umbrales para considerar distintas categorías de ajuste del modelo (como pobre ajuste, aceptable o excelente). No obstante, más allá de esa categorización, para esta tesis el interés se centra sobre cuáles modelos muestran un mejor ajuste, por lo que únicamente se analizará el cambio o la comparación del valor para distintos modelos.

Para efectos de simplicidad en la presentación, seguidamente sólo se muestran tres de los seis escenarios simulados para el nivel inferior o *dentro de grupos*, ya que los resultados fueron similares para los restantes escenarios.

Los gráficos 1, 2 y 3 muestran en términos generales que el uso de una distribución a priori fuertemente informativa en las cargas factoriales cruzadas produce mayores valores del PPP.

En el escenario 1, que presenta las distribuciones a priori más informativas para todos los parámetros del modelo, la diferencia en el ajuste al comparar las distribuciones a priori en las cargas factoriales cruzadas se maximiza, ya que para los casos en los que se utilizan distribuciones difusas o débilmente informativas el ajuste no solo es relativamente bajo, sino similar entre ambos casos. Este hecho se refuerza incluso para las condiciones de pequeños tamaños de muestra, pequeño número de grupos y bajos valores del ICC, lo que significa que cuando las condiciones de simulación son más propicias para realizar estimación frecuentista (gran tamaño de muestra, gran número de grupos, altos valores del ICC) las diferencias se reducen, pero se mantiene la tendencia de mejor ajuste en el caso de distribuciones a priori informativas en las cargas factoriales cruzadas. Sin embargo, entre menos información se incluya en la estimación de los demás parámetros (escenarios 3 y 6), el ajuste bajo las condiciones de distribución a priori difusa en las cargas factoriales cruzadas empeora notoriamente, sin importar otros elementos (como mayor tamaño de muestra).

Bajo estos escenarios, el ajuste al comparar distribuciones a priori en las cargas factoriales cruzadas débilmente informativas y difusas tiende a ser similar, especialmente para grandes muestras y valores del ICC; sin embargo, se refuerza el hecho de que cuando se tienen pocos datos, un pequeño número de grupos y valores bajos del ICC, el uso de distribuciones a priori muy informativas en las cargas factoriales cruzadas sí produce mejores ajustes del modelo global.

Gráfico 1 PPP-Value, escenario 1: Intervalo de credibilidad del P-value Predictivo a Posteriori, según información en la varianza de la distribución a priori de las Cargas Factoriales Cruzadas, Tamaño de la Muestra y Coeficiente de Correlación Intraclass

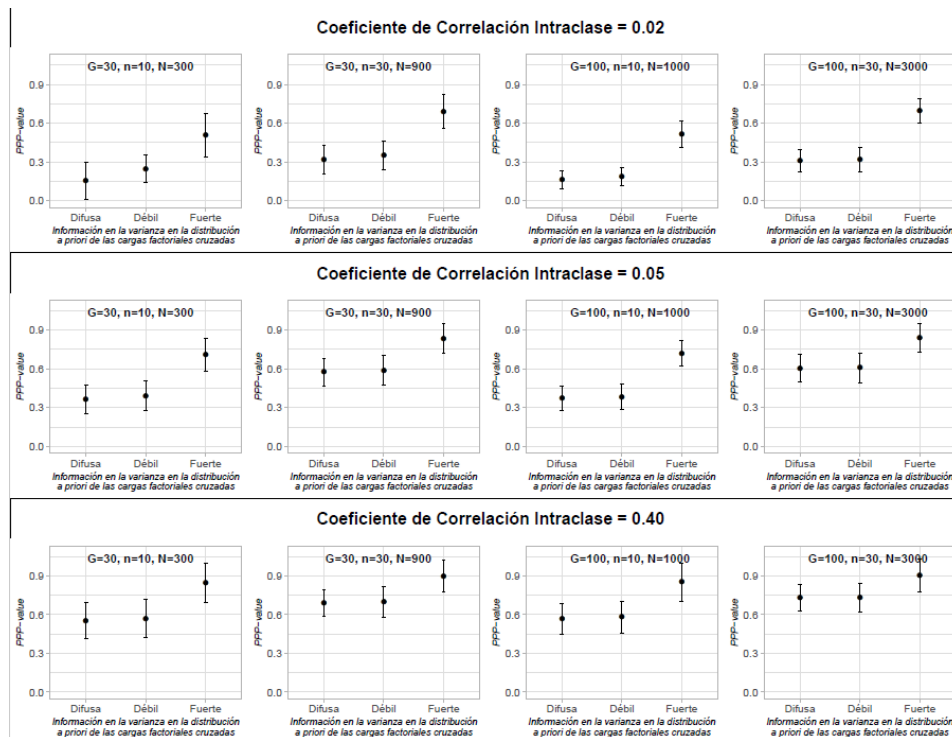


Gráfico 2 PPP-Value, escenario 3: Intervalo de credibilidad del P-value Predictivo a Posteriori, según información en la varianza de la distribución a priori de las Cargas Factoriales Cruzadas, Tamaño de la Muestra y Coeficiente de Correlación Intraclass

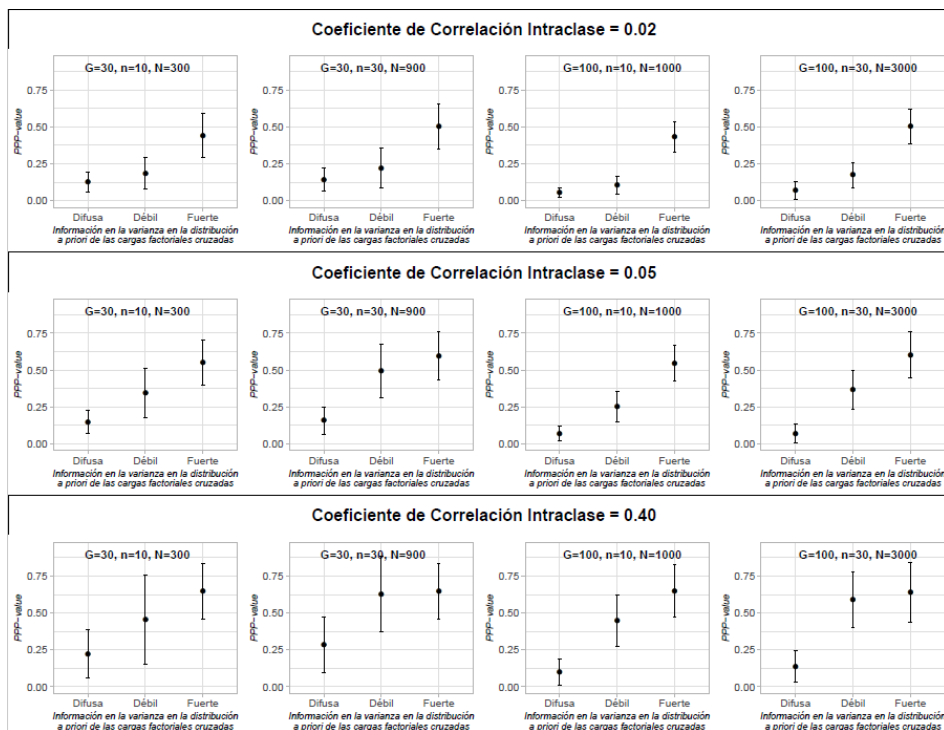
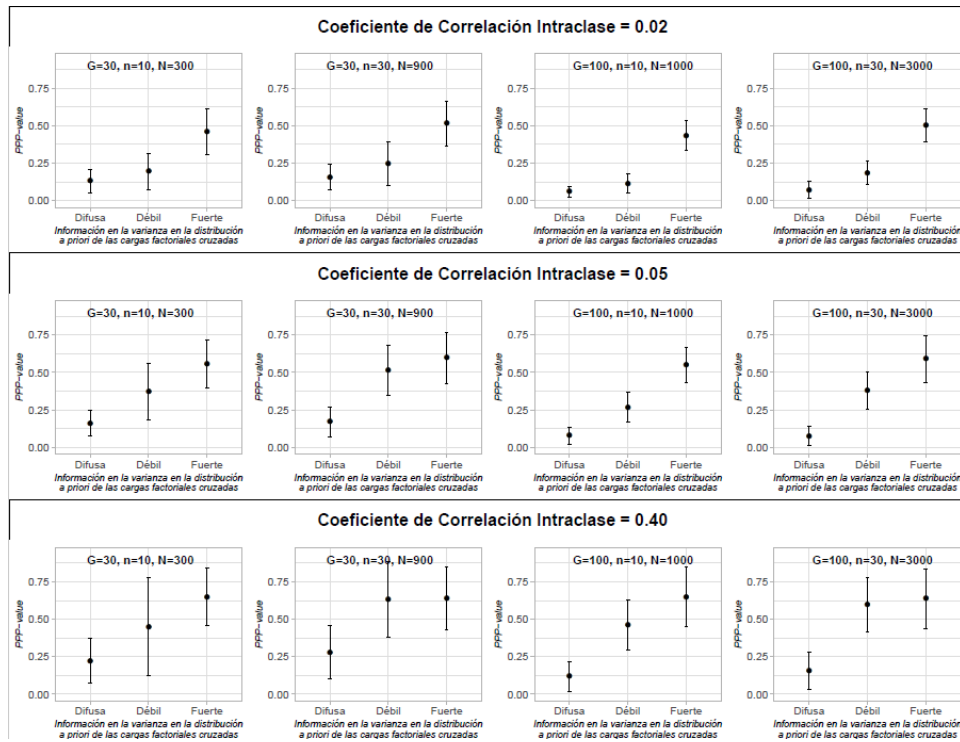


Gráfico 3 **PPP-Value**, escenario 6: Intervalo de credibilidad del P-value Predictivo a Posteriori, según información en la varianza de la distribución a priori de las Cargas Factoriales Cruzadas, Tamaño de la Muestra y Coeficiente de Correlación Intraclass



5.2 Caso aplicado

El caso empírico se analizó siguiendo los lineamientos de la sección 4.2.5, donde se especificó que la relación de distintos factores latentes sobre el rendimiento educativo de los estudiantes costarricenses de 15 años de edad participantes en la evaluación PISA del año 2012. Además de la medición de habilidades y conocimientos de los estudiantes mediante una prueba escrita, PISA también realiza cuestionarios para recolectar información que permita analizar y contextualizar adecuadamente los resultados de las pruebas. Se aplican cuestionarios a los estudiantes (*Student Questionnaire for PISA 2012*) para obtener información sobre aspectos del hogar y la familia. Asimismo, se aplican cuestionarios a los directores de cada centro educativo (*School Questionnaire for PISA 2012*) para obtener información específica sobre los aspectos propios de la infraestructura del centro educativo, cantidad y preparación del personal docente, entre otros aspectos.

Los valores del rendimiento de los estudiantes en las evaluaciones de PISA son considerados como variables latentes continuas, lo que significa que se tratan como cantidades que son medibles pero para las que no existe un instrumento de medida, ya que los ítems no son evaluados con puntuaciones específicas, sino que PISA define escalas a partir de las cuales se realizan las comparaciones. Así, para facilitar la interpretación de las notas asignadas a los estudiantes, las escalas en lectura, matemáticas y ciencias fueron diseñadas para tener una nota promedio de 500 puntos y una desviación estándar de 100 puntos para los países miembros de la OECD. Esto significa que cerca de dos tercios de los estudiantes de países miembros de la OECD obtienen notas entre los 400 y 600 puntos (OECD, 2009). A estos valores del rendimiento de los estudiantes se les llama valores plausibles.

Como ha sido explicado, la modelación multinivel es necesaria cuando existe dependencia de las unidades del primer nivel en relación con los conglomerados en los que se agrupan. En los datos analizados para Costa Rica, es posible calcular el coeficiente de correlación intraclase (CCI). Este es de 0,468, el cual es un resultado bastante alto, y evidenciando una alta dependencia de la variabilidad de los resultados académicos de acuerdo con la agrupación en los centros educativos. Este resultado muestra las grandes desigualdades que existen en los centros educativos del país y que terminan incidiendo en la distribución de los resultados educativos.

Fueron modeladas relaciones tanto en el nivel inferior como superior. Es decir, en el nivel inferior fue modelada la relación de cinco factores latentes recolectados a nivel del estudiante sobre el resultado individual de cada uno de los estudiantes en la evaluación de alfabetización matemática, mientras que en el nivel superior fue modelada la relación de tres factores latentes recolectados a nivel de centro educativo sobre el rendimiento medio del centro educativo.

El objetivo de la modelación del caso aplicado mediante modelos de ecuaciones estructurales consiste en evaluar la consistencia interna del modelo hipotetizado, es decir, si los constructos definidos realmente capturan el factor latente que se intenta modelar, así como la relación entre los distintos factores sobre el rendimiento educativo.

Pero más allá de ese resultado, los hallazgos obtenidos en esta investigación sugieren que la estimación del modelo SEM multinivel mediante métodos Bayesianos y permitiendo que algunas cargas factoriales cruzadas sean estimadas libremente produce ganancias en términos del ajuste del

modelo particularmente, sin detrimento de otros elementos de interés como el valor puntual de las estimaciones de las cargas, las varianzas y otros parámetros del modelo.

Por lo anterior, el modelo planteado fue estimado bajo dos escenarios: el primer escenario corresponde al enfoque tradicional, de delimitar las cargas factoriales cruzadas a ser exactamente iguales a cero (parámetros fijos), mientras que el segundo escenario corresponderá al modelo en el cual se introducirá la estimación de cargas factoriales cruzadas con una distribución a priori distribuida normalmente con media cero y una varianza muy pequeña, siguiendo el enfoque BSEM explicado en la sección 2.5.

Seguidamente se explicarán los resultados obtenidos analizando cada sección del modelo estimado por separado, pero comparando simultáneamente los dos escenarios descritos en el párrafo anterior.

En cuanto al ajuste del modelo, el indicador utilizado es el PPP. Si bien aún no se cuenta con la suficiente evidencia empírica para establecer valores de corte para definir lo que puede ser considerado como buen ajuste, entre mayor sea su valor (varía entre cero y uno) mejor es el ajuste del modelo.

En el caso del escenario 1, sin liberación en la estimación de las cargas factoriales cruzadas, el PPP obtenido fue de 0.271, mientras que para el escenario 2 el PPP se incrementa hasta 0.334.

El resultado anterior efectivamente muestra un incremento en la bondad del ajuste global del modelo. El escenario 2 incluye cargas factoriales cruzadas con varianzas pequeñas en la distribución a priori de todos los factores latentes independientes del modelo (cinco en el nivel dentro de grupos, tres en el nivel entre grupos).

Al evaluar las cargas factoriales en el nivel dentro de grupos, se obtienen los resultados que se muestran en la Tabla 27. Lo más destacado de dicha tabla son las altas cargas factoriales que se muestran en cuatro de los cinco constructos, con excepción del constructo de antecedentes educativos.

Todos los coeficientes se muestran en valores estandarizados, por lo que estas cargas factoriales superiores a 0.70 en valor absoluto muestran que las relaciones hipotetizadas y planteadas en conjunto con expertos en el área sí se encuentran acorde con la estructura de datos observada en PISA.

En el caso del constructo de antecedentes educativos, que incluye las variables indicadores de asistencia a educación preescolar y de repitencia en educación primaria o secundaria, los resultados indican que ambas variables no corresponden al mismo constructo, es decir, que no aportan variabilidad sobre el mismo factor latente, lo cual no era esperable al asumir que ambas variables, al medir hechos ocurridos previamente en el historial del estudiante, podrían corresponder a un constructo similar.

Tabla 27 Comparación de las estimaciones de las cargas factoriales de los factores latentes del modelo dentro de grupos

Factor latente	Variable Indicadora	Escenario 1				Escenario 2			
		Estimación	Posterior SD	P-value	Sig	Estimación	Posterior SD	P-value	Sig
Antecedentes educativos	ANT1	1,000	0,000	0,000		1,000	0,000	0,000	
	ANT2	-0,160	0,045	0,000	***	0,056	0,072	0,209	
Extracción socioeconómica	EC1	1,000	0,000	0,000		1,000	0,000	0,000	
	EC2	0,720	0,021	0,000	***	0,788	0,035	0,000	***
	EC3	0,762	0,023	0,000	***	0,735	0,042	0,000	***
	EC4	0,839	0,025	0,000	***	0,781	0,044	0,000	***
Actitud personal hacia las matemáticas	ACTP1	1,000	0,000	0,000		1,000	0,000	0,000	
	ACTP2	-0,812	0,044	0,000	***	-0,856	0,064	0,000	***
	ACTP3	0,906	0,053	0,000	***	0,798	0,073	0,000	***
	ACTP4	0,694	0,050	0,000	***	0,831	0,071	0,000	***
	ACTP5	0,764	0,055	0,000	***	0,798	0,075	0,000	***
Actitud hacia el entorno	ACTEN1	1,000	0,000	0,000		1,000	0,000	0,000	
	ACTEN2	0,746	0,066	0,000	***	0,792	0,072	0,000	***
	ACTEN3	0,816	0,081	0,000	***	0,836	0,106	0,000	***
	ACTEN4	0,862	0,081	0,000	***	0,896	0,094	0,000	***
Conocimientos matemáticos	FAMCON	1,000	0,000	0,000		1,000	0,000	0,000	
Alfabetización matemática	PV1	1,000	0,000	0,000		1,000	0,000	0,000	
	PV2	0,990	0,009	0,000	***	0,990	0,009	0,000	***
	PV3	0,961	0,010	0,000	***	0,961	0,010	0,000	***
	PV4	0,960	0,010	0,000	***	0,960	0,010	0,000	***
	PV5	0,988	0,010	0,000	***	0,988	0,010	0,000	***
	PV6	0,932	0,010	0,000	***	0,933	0,011	0,000	***
	PV7	0,943	0,010	0,000	***	0,944	0,010	0,000	***

Posterior SD: Desviación estándar del intervalo de credibilidad de las estimaciones a posteriori

Sig: Nivel de significancia al 1% ***, 5%*.

Tabla 28 Cargas factoriales cruzadas con estimaciones distintas del parámetro cero (Escenario 2)

Factor latente	Variable Indicadora	Escenario 2			
		Estimación	Posterior SD	P-value	Sig
Extracción Socioeconómica	ACTP1	0,171	0,041	0,000	***
	FAMCON	0,203	0,072	0,000	***
Actitud Personal hacia el centro educativo	FAMCON	0,160	0,073	0,000	***
Actitud hacia el entorno	ANT1	0,215	0,086	0,000	***
	ACTP4	0,131	0,067	0,000	***
	FAMCON	0,362	0,077	0,000	***
Conocimientos matemáticos	ANT1	-0,256	0,073	0,000	***
	ANT2	-0,129	0,055	0,000	***
	EC2	0,122	0,071	0,000	***
	ACTP2	-0,141	0,068	0,000	***
	ACTP4	-0,197	0,064	0,000	***
	ACTP5	-0,135	0,066	0,000	***

La tabla 28 es un complemento a la tabla 27. Esta muestra la estimación final resultante de la liberación de parámetros que en el escenario 1 han sido fijados como cero, pero que en el escenario 2 se han modelado con una distribución a priori que los define como parámetros aproximadamente iguales a cero.

Lo esperable era que todas las cargas factoriales cruzadas tuvieran estimaciones cercanas a cero o cuyo intervalo de credibilidad incluyera el cero. Al descartar todos los casos con estimaciones inferiores al valor estimado de 0.10, se obtiene lo que se observa en la tabla 28.

Esta tabla muestra que, a pesar de que teóricamente la relación entre esas variables indicadoras y el constructo latente era nula (no se hipotetizaba relación alguna), dados los datos observados sí se halló evidencia de algún tipo de relación débil.

Esto sucedió particularmente en el caso del constructo de conocimientos matemáticos, que muestra un alto grado de asociación con las variables indicadores del constructo de antecedentes educativos y de actitud personal hacia las matemáticas.

Esto plantea un reto adicional para la aplicación del enfoque BSEM, ya que la evidencia muestra que las relaciones planteadas muestran una ligera desviación con respecto a lo indicado por el investigador.

Finalmente, en cuanto a las relaciones estructurales planteadas en el modelo, de relación directa entre los cinco constructos y la alfabetización matemática, se muestran altos valores de los parámetros de regresión para los constructos de extracción socioeconómica, actitud personal hacia las matemáticas, actitud hacia el entorno y conocimientos matemáticos. En el caso de los antecedentes educativos, la relación es más débil con respecto a los demás constructos.

Tabla 29 Estimación de los parámetros estructurales del nivel dentro de grupos

Parámetro	Escenario 1				Escenario 2			
	Estimación	Posterior SD	P-value	Sig	Estimación	Posterior SD	P-value	Sig
Antecedentes educativos	0,071	0,020	0,000	***	0,031	0,025	0,000	***
Extracción socioeconómica	0,358	0,016	0,000	***	0,398	0,020	0,000	***
Actitud personal hacia matemáticas	0,303	0,013	0,000	***	0,337	0,017	0,000	***
Actitud hacia el entorno	0,284	0,019	0,000	***	0,316	0,024	0,000	***
Conocimientos matemáticos	0,242	0,017	0,000	***	0,269	0,021	0,000	***

Para el modelo superior (entre grupos) la tabla 28 muestra altas cargas factoriales en los tres constructos definidos para este modelo, con una tendencia a mayores cargas en el caso del escenario 2.

Tabla 30 Comparación de las estimaciones de las cargas factoriales de los constructos del modelo entre grupos

Factor latente	Variable Indicadora	Escenario 1				Escenario 2			
		Estimación	Posterior SD	P-value	Sig	Estimación	Posterior SD	P-value	Sig
Calidad recursos educativos	TCHORT	1,000	0,000	0,000		1,000	0,000	0,000	
	SCMATEDU	0,935	0,236	0,000	***	0,956	0,208	0,000	***
	SCMATBUI	0,905	0,226	0,000	***	0,946	0,202	0,000	***
Clima escolar	STUDCLIM	1,000	0,000	0,000		1,000	0,000	0,000	
	TEACCLIM	0,763	0,114	0,000	***	0,882	0,095	0,000	***
Calidad docente	TCMORALE	1,000	0,000	0,000		1,000	0,000	0,000	
	TCFOCST	0,773	0,175	0,000	***	0,668	0,113	0,000	***
Alfabetización matemática	PVSCH1	1,000	0,000	0,000		1,000	0,000	0,000	
	PVSCH2	0,980	0,022	0,000	***	0,981	0,022	0,000	***
	PVSCH3	0,933	0,024	0,000	***	0,932	0,024	0,000	***
	PVSCH4	0,915	0,027	0,000	***	0,914	0,028	0,000	***
	PVSCH5	0,919	0,029	0,000	***	0,919	0,029	0,000	***
	PVSCH6	0,969	0,017	0,000	***	0,970	0,018	0,000	***
	PVSCH7	0,943	0,027	0,000	***	0,941	0,027	0,000	***

En cuanto a las cargas factoriales cruzadas, únicamente en uno de los casos se obtuvo una estimación superior a 0.10 en valor absoluto, en el constructor de clima escolar con la carga cruzada de escasez de personal docente (del constructo de calidad en los recursos educativos).

Los parámetros estructurales, por su parte, muestran también relaciones altas en la estimación de los parámetros de regresión, superiores a 0.30 en los tres constructos de este nivel.

6 Conclusiones

Siguiendo la metodología propuesta por Muthén & Asparouhov (2012), esta investigación incursionó en una discusión reciente en la comunidad científica en el contexto de modelos de ecuaciones estructurales Bayesianos multinivel (ML-BSEM).

En la última década se ha discutido el uso de varianzas muy pequeñas en las distribuciones a priori de las cargas factoriales cruzadas en el contexto de la estimación de modelos de análisis factorial confirmatorio o de ecuaciones estructurales con enfoque Bayesiano.

Tradicionalmente, las investigaciones que incursionan en el uso del análisis factorial confirmatorio establecen las cargas factoriales cruzadas como iguales a cero, queriendo indicar con esto que el indicador particular no tiene ninguna relación con el factor latente o constructo para el cual la carga factorial ha sido establecida como exactamente igual a cero.

Sin embargo, las variables indicadoras rara vez son indicadores de construcción perfectamente puros, especialmente en el estudio de las ciencias sociales, lo que implica que se pueden presentar niveles significativos de asociación con múltiples constructos.

Por tal razón, en esta tesis se abordó este tema al plantear modelos en los que fue liberada la estimación de parámetros en las cargas factoriales cruzadas. Este nuevo enfoque busca mantener los pilares básicos de la investigación no exploratoria, en el sentido de que no se trata de ajustar las relaciones hipotetizadas por el investigador para obtener mejoras en la estimación del modelo, sino que lo que se busca es aprovechar las ventajas de la estimación Bayesiana para mejorar el ajuste del modelo sin modificar las hipótesis del investigador, al variar la estimación de parámetros de ceros exactos por estimación de parámetros aproximadamente iguales a cero, mediante el uso de distribuciones a priori con varianzas sumamente pequeñas en la estimación de dichos parámetros.

Los resultados obtenidos en esta investigación refuerzan el hecho ampliamente comprobado de que la estimación Bayesiana puede ser utilizada para obtener estimaciones admisibles en situaciones en que la estimación frecuentista tiende a fallar. Si bien no fueron estimados modelos mediante métodos frecuentistas, todos los modelos estimados en esta tesis con enfoque Bayesiano convergieron sin importar el número de grupos en la estructura jerárquica, el tamaño de los grupos o la fuerza del coeficiente de correlación intraclase utilizada.

El objetivo general de esta investigación era el de comparar el desempeño de modelos de ecuaciones estructurales Bayesianos multinivel según distintos escenarios de estimación de las cargas factoriales cruzadas. Para hacer esto, fueron establecidos tres niveles de información a introducir en la varianza de las cargas factoriales cruzadas (poco informativa, débilmente informativa e informativa), siendo uno de estos el de interés para validar el enfoque propuesto por los Muthén & Asparouhov (2012), de utilizar una distribución a priori con varianza muy pequeña o sumamente informativa.

La comprobación de la hipótesis en un contexto más amplio requirió la generación de un estudio de simulación con distintos factores y niveles. El estudio permitió mostrar que, en el nivel inferior (o *dentro de grupos*) el uso de distribuciones a priori informativas en las cargas factoriales cruzadas

mejora la precisión con la que son estimados los parámetros principales (cargas factoriales directas y parámetros estructurales), ya que el intervalo de credibilidad al 95% de estas estimaciones se reduce consistentemente al aumentar la cantidad de información en las varianzas cruzadas.

Estas mejoras son incluso más pronunciadas cuando se toma en cuenta el tamaño de la muestra, ya que entre mayor número de datos menor es el intervalo de credibilidad del parámetro, especialmente en los casos con gran número de grupos. Además, entre menos informativas sean las distribuciones a priori del resto de parámetros de interés, también mayor amplitud tendrán los intervalos de credibilidad.

El RMSE mostró altos valores en la mayoría de parámetros de interés cuando la distribución a priori de la carga factorial cruzada es difusa. En general, el RMSE cuando la distribución a priori de la carga factorial cruzada es débilmente informativa o informativa presenta un buen comportamiento (valores bajos).

En cuanto a los niveles de recuperación del parámetro, estos fueron consistentemente altos (superiores a 0.95) en la mayoría de escenarios simulados para el parámetro de cargas factoriales directas. Pero para las cargas factoriales cruzadas y el parámetro estructural, únicamente los casos con distribución a priori informativa en la carga factorial cruzada muestran niveles aceptables de recuperación del parámetro en todos los escenarios. El parámetro de varianza residual y factorial no logró obtener niveles aceptables de recuperación del parámetro para los escenarios simulados.

En el nivel superior jerárquico o *entre grupos*, a nivel general la estimación Bayesiana con varianzas a priori informativas en las cargas factoriales cruzadas presenta resultados muy superiores a los de las demás condiciones de estimación en el nivel entre grupos. Sin embargo, sí existen variaciones en cuanto al nivel de sesgo dependiendo del tipo de parámetro.

Al analizar el RMSE de las estimaciones en este nivel, a lo largo de todas las condiciones que fueron simuladas, las estimaciones Bayesianas con distribuciones a priori débilmente informativas e informativas en las cargas factoriales cruzadas mostraron valores pequeños del RMSE consistentemente. Para la estimación Bayesiana con distribución a priori difusa, el RMSE estuvo inversamente asociado al número de grupos, el tamaño de los grupos y el ICC.

Finalmente, en cuanto al ajuste global del modelo, y mediante el uso del PPP-Value, el ajuste al comparar distribuciones a priori en las cargas factoriales cruzadas débilmente informativas y difusas tiende a ser similar cuando se cuenta un alto grado de información en la estimación de otros parámetros del modelo, especialmente para grandes muestras y valores del ICC; sin embargo, se refuerza el hecho de que cuando se tienen pocos datos, un pequeño número de grupos y valores bajos del ICC, el uso de distribuciones a priori muy informativas en las cargas factoriales cruzadas sí produce mejores ajustes del modelo global.

Se puede afirmar que, de forma consistente, el uso de distribuciones a priori con varianza pequeña o muy informativa en las cargas factoriales cruzadas produce mejores resultados bajo distintas condiciones de estimación de los modelos de ecuaciones estructurales multinivel, por lo que esta investigación aporta evidencia para que este enfoque pueda ser replicada en otras investigaciones que busquen obtener ganancias en la estimación de modelos de ecuaciones estructurales Bayesianos multinivel. Sin embargo, hasta que se hayan realizado más investigaciones para establecer métodos adecuados en las ciencias sociales para determinar el tamaño exacto de los

hiperparámetros de varianza en las distintas distribuciones a priori, es recomendable que los investigadores interesados en un enfoque bayesiano de SEM multinivel siempre realicen un análisis de sensibilidad para determinar en qué medida las estimaciones del modelo dependen de supuestos previos.

7 Referencias

- Albert, J. (2009). *Bayesian computation with R*. Springer. Second edition.
- Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis of latent variable models using Mplus (Mplus Technical Report)*. Retrieved from statmodel.com/download/BayesAdvantages18.pdf.
- Asparouhov, T., & Muthén, B. (2012). *Bayesian exploratory factor analysis*. No publicado.
- Asparouhov, T., & Muthén, B. (2019). *Bayes Parallel Computation: Choosing the number of processors*. Mplus project.
- Asparouhov, T., Muthén, B., & Morin, A. (2015). *Bayesian Structural Equation Modeling With Cross-Loadings and Residual Covariances: Comments on Stromeyer et al.* *Journal of Management*, Vol. XX No. X, Month XXXX 1–17.
- Babenko, O., Alves, C., & Bahry, L. (2012). *Using Structural Equation Modeling to Investigate Students' Career Awareness in Science*. *Canadian Journal for New Scholars in Education*. Volume 4, Issue 1.
- Banerjee, A., Cole, S., & Duflo, E. (2007). *Remedying Education: Evidence from Two Randomized Experiments in India*. *Quarterly Journal of Economics* 122 (3): 1235-64.
- Barrera-Osorio, F., & Linden, L. (2009). *The Use and Misuse of Computers in Education: Evidence from a Randomized Experiment in Colombia*. Policy Research Working Paper Series. Washington, DC: World Bank.
- Bauer, D. J. (2003). *Estimating multilevel linear models as structural models*. *Journal of Educational and Behavioral Statistics*, 28, 135–167.
- Bentler, P. M., & Liang, J. (2003). *Two-level mean and covariance structures: Maximum likelihood via and EM algorithm*. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 53–70). Hillsdale, NJ: Erlbaum.
- Beuermann, D., Cristia, D., Cruz-Aguayo, J., Cueto, S., & Malamud, O. (2015). *Home computers and child outcomes: Short-term impacts from a randomized experiment in Peru*. *American Economic Journal: Applied Economics*, 7(2), 53-80.
- Brooks, S., & Gelman, A. (1998). *General methods for monitoring convergence of iterative simulations*. *Journal of Computational and Graphical Statistics*, 7, 434–455.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: the Guilford Press.

- Coleman, J. (1966). *Equality of Educational Opportunity*. Washington D.C.: UD Department of Education.
- Cristia, J. (2012). *Technology and child development: evidence from the one laptop per child program*. IDB Working Paper Series, N° 304, Washington, D.C., Department of Research and Chief Economist, Banco Interamericano de Desarrollo.
- Cristia, J., Czerwonko, A., & Garofalo, P. (2010). *Does ict Increase Years of Education? Evidence from Peru*. ILO Working Papers, N° 110, Washington, D.C., Banco Interamericano de Desarrollo.
- Curran, P. J. (2003). *Have multilevel models been structural equation models all along?* *Multivariate Behavioral Research*, 38, 529–569.
- Depaoli, S., & Clifton, J. (2015). *A bayesian approach to Multilevel Structural Equation Modeling with Continuous and Dichotomous Outcomes*. *Structural Equation Modeling: A multidisciplinary Journal*.
- Díaz Porras, R., & Jiménez Rodríguez, R. (1980). *Consideraciones económicas sobre el sistema educativo costarricense*. Tesis para optar al grado de licenciatura, San José, Facultad de Ciencias Económicas, Universidad de Costa Rica.
- Diya, L., Li, B., Heede, K., Sermeus, W., & Lesaffre, E. (2013). *Multilevel factor analytic models for assessing the relationship between nurse-reported adverse events and patient safety*. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.
- du Toit, S., & du Toit, M. (2003). *Multilevel structural equation modeling*. In J. De Leeuw & I. G. G. Kreft (Eds.), *Handbook of quantitative multilevel analysis* (pp. 273–321). Boston: Kluwer.
- Duflo, E., & Banerjee, A. (2017). *Handbook of Field Experiments*. Volume 1. North Holland.
- Fernández Aráuz, A. (2013). *Análisis de la Resiliencia Educativa de los estudiantes costarricenses con datos de la Prueba de Lectura de la Evaluación PISA 2009*. *Revista de Ciencias Económicas* Vol 31. N°2. ISSN: 0252-9521. Universidad de Costa Rica.
- Fernández Aráuz, A. (2014). *La influencia del acceso al uso de computadoras en edad temprana sobre el rendimiento en matemáticas*. *Revista de Ciencias Económicas* 32-No.2: 2014 / 113-125 / ISSN: 0252-9521. Universidad de Costa Rica.
- Fernández Aráuz, A., & Del Valle Alvarado, R. (2013). *Desigualdad Educativa en Costa Rica: La brecha entre los estudiantes de colegios públicos y privados*. *Revista CEPAL* N°111.
- Fernández, A. (2017). *Un modelo de ecuaciones estructurales bayesiano: aplicación al rendimiento matemático en PISA 2012*. *Revista Estadística Española*. Volumen 59, número 192 / 2017, 5-27.

- Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel Modeling using R*. Chapman & Hall/CRC statistics in the social and behavioral sciences.
- Finch, W., & French, B. (2011). *Estimation of MIMIC model parameters with multilevel data*. *Structural Equation Modeling*, 18, 229–252.
- Geman, S., & Geman, D. (1984). *Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Geyer, C. (1992). *Practical Markov chain Monte Carlo*. *Statistical Science*, 7, 473–511.
- Giménez, G., Arias R, R., Castro A, G., Fernández, L. J., Ospina P, N., & Sánchez, L. (2014). *¿Por qué los estudiantes de colegios públicos y privados de Costa Rica obtienen distintos resultados académicos?* Quinto Informe del Estado de la Educación. Programa Estado de la Nación en Desarrollo Humano Sostenible.
- Glewwe, P., Hanushek, E., Humpage, S., & Ravina, R. (2011). *School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010*. NBER Working Papers 17554, National Bureau of Economic Research, Inc. DOI: 10.7208/chicago/9780226078854.001.0001.
- Glewwe, P., Kremer, M., & Moulin, S. (2009). *Many Children Left Behind? Textbooks and Test Scores in Kenya*. *American Economic Journal: Applied Economics* 1 (1): 112-35.
- Goldstein, H., Bonnet, G., & Rocher, T. (2007). *Multilevel Structural Equation Models for the Analysis of Comparative Data on Educational Performance*. *Journal of Educational and Behavioral Statistics*. Vol. 32, No. 3, pp. 252–286.
- Hallquist, M., & Wiley, J. (2018). *MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus*. *Structural Equation Modeling*, 25, 621-638. doi: 10.1080/10705511.2017.1402334.
- Hanushek, E. (1986). *The economics of Shooling*. *Journal of Economic Literature* 24(3), 1141-77.
- Hoijtink, H., & van de Schoot, R. (2017). *Testing Small Variance Priors Using Prior-Posterior Predictive P-values*. Department of Methodology and Statistics, Utrecht University.
- Holzinger, K., & Swineford, F. (1939). *A study in factor analysis: The stability of a bifactor solution*. *Supplementary educational monographs*. Chicago, IL: University of Chicago.
- Hox, J. J., & Maas, C. J. (2001). *The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples*. *Structural Equation Modeling*, 8, 157–174. doi:10.1207/S15328007SEM0802_1.

- James, L., Mulaik, S., & Brett, J. (1982). *Causal analysis: Assumptions, models and data*. Beverly Hills, CA: Sage.
- Joreskog, K. (1969). *A general approach to confirmatory maximum likelihood factor analysis*. *Psychometrika*, 34, 183-202.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8. Structural Equation Modeling with the SIMPLIS Command Language*. Chicago. Lawrence Erlbaum Associates Publishers.
- Kaplan, D. (2009). *Structural equation modeling: foundations and extensions*. Segunda edición. Thousand Oaks, CA: Sage.
- Kaplan, D., & Su, D. (2015). *Context Questionnaire Rotation and Imputation with Implications for Estimation of Plausible Values in Large-Scale Assessments*. University of Connecticut, Connecticut, United States.: Modern Modeling Methods Conference 2015.
- Kaplan, D., Kim, J., & Kim, S. (2009). *Multilevel latent variable modeling: Current research and recent developments*. In R. E. Millsap & A. Maydeu-Olivares (Eds.). *The Sage handbook of quantitative methods in psychology* (pp. 592–612). Thousand Oaks, CA: Sage.
- Kass, R., & Raftery, A. (1995). *Bayes factors*. *Journal of the American Statistical Association*, 90, 773–795.
- Kim, M., & Song, J. (2010). *A Confirmatory Structural Equation Model of Achievement Estimated by Dichotomous Attitudes, Interest, and Conceptual Understanding*. *Eurasia Journal of Mathematics, Science & Technology Education*, 2010, 6(4), 271-285.
- Lee, S. Y. (2007). *Structural Equation Modeling: A Bayesian Approach*. Wiley Series in probability and statistics.
- Lee, S. Y., & Shi, J. Q. (2001). *Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data*. *Biometrics*, 57, 787–794.
- Lee, S., & Song, X. (2001). *Hypothesis Testing and Model Comparison in Two-level Structural Equation Models*. *Multivariate Behavioral Research*, 36 (4), 639-655. The Chinese University of Hong Kong.
- Lee, S., Poon, W., & Bentler, P. (1990). *Full maximum likelihood analysis of structural equation models with polytomous variables*. *Statistics and Probability Letters*, 9, 91–97.
- Li, X., & Beretvas, S. (2013). *Sample size limits for estimating upper level mediation models using multilevel SEM*. *Structural Equation Modeling*, 20, 241–264.
doi:10.1080/10705511.2013.769391.
- Little, J. (2013). *Multilevel confirmatory ordinal factor analysis of the life skills profile*. *Psychological Assessment*. Advance online publication.

- Little, T., & Card, N. (2013). *Longitudinal Structural Equation Modeling (Methodology in the Social Sciences)*. New York: Guilford Press. March 25, 2013.
- Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). *A 2x2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models*. *Psychological Methods*, 16, 444–467. doi:10.1037/a0024376.
- Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). *The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies*. *Psychological Methods*, 13, 203–229. doi:10.1037/a0012869.
- MacCallum, R., & Edwards, M. (2012). *Hopes and Cautions in Implementing Bayesian Structural Equation Modeling*. *Psychological Methods*. Vol. 17, No. 3, 340–345.
- Malamud, O., & Pop-Eleches, C. (2011). *Home computer use and the development of human capital*. *The Quarterly Journal of Economics*, 126, 987-1027.
- MEP. (2010). *Informe Nacional Pruebas Nacionales Diagnósticas de II Ciclo de la Educación General Básica, 2008*. San José, Costa Rica: Dirección de Gestión y Evaluación de la Calidad, Departamento de Evaluación Académica y Certificación, Ministerio de Educación Pública.
- MEP. (2012). *Informe Nacional de Factores Asociados al Rendimiento Académico en las Pruebas Nacionales Diagnósticas, III Ciclo de la Educación General Básica, 2010*. San José, Costa Rica: Dirección de Gestión y Evaluación de la Calidad, Departamento de Evaluación Académica y Certificación, Ministerio de Educación Pública.
- Metha, P., & Neale, M. (2005). *People are variables too: Multilevel structural equations modeling*. *Psychological Methods*, 10(3), 259-284. doi: 10.1037/1082-989x.10.3.259.
- Montero, E. (2012). *Los modelos de ecuaciones estructurales como herramienta para explorar posibles relaciones causales en investigación educativa: Una ilustración con datos de PISA 2009 en Costa Rica*. Aporte Especial. Cuatro Informe del Estado de la Educación. Programa Estado de la Nación en Desarrollo Humano Sostenible.
- Montero, E., Rojas, S., & Zamora, E. (2014). *Costa Rica En Las Pruebas Pisa 2012*. Quinto Informe del Estado de la Educación. Programa del Estado de la Nación en Desarrollo Humano Sostenible.
- Montero, E., Rojas, S., Zamora, E., & Rodino, A. (2013). *Costa Rica en las pruebas PISA 2009 de Competencia Lectora y Alfabetización Matemática*. Cuarto Informe del Estado de la Educación: Programa Estado de La Nación en Desarrollo Humano Sostenible.
- Moreira Mora, T. (2009). *Factores endógenos y exógenos asociados al rendimiento en matemática: Un análisis multinivel*. *Revista Educación*, vol. 33, N° 2, San José, Universidad de Costa Rica.

- Morselli, D., Spini, D., & Devos, T. (2012). *Human values and trust in institutions across countries: A multilevel test of Schwartz's hypothesis of structural equivalence*. *Survey Research Methods*, 6, 49–60.
- Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. New York: CRC Press Taylor & Francis Group.
- Muralidharan, K., & Sundararaman, V. (2008). *Contract Teachers: Experimental Evidence from India*. JPAL Working Paper, MIT.
- Muthén, B. (1989). *Latent variable modeling in heterogeneous populations*. *Psychometrika*, 54, 557–585.
- Muthén, B. (1991). *Multilevel factor analysis of class and student achievement components*. *Journal of Educational Measurement*, 28, 338–354.
- Muthén, B. (1994). *Multilevel covariance structure analysis*. *Sociological Methods & Research*, 22, 376–398.
- Muthén, B., & Asparouhov, T. (2012 (1)). *Bayesian SEM: A more representation of substantive theory*. *Psychological Methods*, 17, 313–335.
- Muthén, B., & Asparouhov, T. (2012 (2)). *Rejoinder to MacCallum, Edwards, and Cai (2012) and Rindskopf (2012): Mastering a New Method*. *Psychological Methods*. Vol. 17, No. 3, 346–353.
- Muthén, B., & Asparouhov, T. (2013(1)). *BSEM Measurement Invariance Analysis*. Mplus Web Notes: No. 17. January 11, 2013. Los Angeles: Mutén & Muthén. www.statmodel.com.
- Muthén, B., & Asparouhov, T. (2013(2)). *New Methods for the Study of Measurement Invariance with Many Groups*. www.statmodel.com.
- Muthén, L., & Muthén, B. ((1998-2015)). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
- Oaxaca, R. (1973). *Male-Female Wage Differentials in Urban Labor Markets*. *International Economic Review*, 14(3), 693–709.
- OECD. (2009). *PISA Data Analysis Manual, SPSS Second Edition*. OECD Publishing.
- OECD. (2012). *PISA Technical Report*. OECD Publishing.
- OECD. (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing.
<http://dx.doi.org/10.1787/9789264190511-en>.

- Olson, U. (1979). *Maximum likelihood estimation of the polychoric correlation coefficient*. *Psychometrika*, 44, 443–460.
- Oviedo, Y. (2012). *Factores asociados al rendimiento académico en Matemática en el III ciclo de la Educación General Básica; Un estudio multinivel*. Estado de la Nación. Costa Rica: CONARE.
- Preacher, K., Zyphur, M., & Zhang, Z. (2010). *A general multilevel SEM framework for assessing multilevel mediation*. *multilevel mediation*.
- Press, S. J. (2003). *Subjective and objective bayesian statistics*. Wiley Series in Probability and Statistics. Second edition.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). *Generalized multilevel structural equation modelling*. *Psychometrika* 69, 167-190.
- Rabe-Hesketh, S., Skrondal, A., & Zheng, X. (2012). *Multilevel structural equation modeling*. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 512–531). New York, NY: Guilford.
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage. (2nd ed).
- Rindskopf, D. (2012). *Next steps in Bayesian structural equation models: Comments on, variations of, and extensions to Muthen and Asparouhov (2012)*. *Psychological Methods*, 17(3), 336–339.
- Rojas Torres, L. (2014). *Construcción y validación del índice de calidad de la gestión de la educación secundaria mediante el análisis factorial confirmatorio*. Tesis para optar el grado de Maestría Académica en Estadística; Universidad de Costa Rica.
- Rojas, L. (2004). *Factores Asociados a la Repitencia de los y las Estudiantes que Cursan Séptimo Año en Colegios Académicos, Diurnos y Públicos: Un Análisis de Niveles Múltiples*. Tesis doctoral. San José, Costa Rica: Universidad Estatal a Distancia.
- Rovine, M. J., & Molenaar, P. C. P. C. (2000). *A structural modeling approach to a multilevel random coefficients model*. *Multivariate Behavioral Research*, 35, 51–88.
- Salas, O. (2012). *Constructo “Alfabetización Matemática”, según PISA*. Cuarto Informe del Estado de la Educación. Programa Estado de la Nación en Desarrollo Humano Sostenible.
- Song, X.-Y., & Lee, S.-Y. (2012). *Basic and Advanced Bayesian Structural Equation Modeling*. Wiley Series in probability and statistics.
- Stromeyer, W., Miller, J., Sriramachandramurthy, R., & DeMartino, R. (2015). *The prowess and pitfalls of Bayesian structural equation modeling: Important considerations for management research*. *Journal of Management*, 41: 491-520.

- Tan, J.-P., Lane, J., & Lassibille, G. (1999). *Student Outcomes in Philippine Elementary Schools: An Evaluation of Four Experiments*. *World Bank Economic Review* 13 (3): 493-508.
- Thomas, A., & OHara, R. (2004). *Openbugs*.
- Urzúa, S., & Veramendi, G. (2011). *The Impact of Out-of-Home Child Care Centers on Early Childhood Development*. Department of Research and Chief Economist, Working Paper Series N°. IDB-WP-240. Inter-American Development Bank.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). *mice: Multivariate Imputation by Chained Equations in R*. *Journal of Statistical Software*, 45(3), 1-67. URL <http://www.jstatsoft.org/v45/i03/>.
- Vegas, E., & Petrow, J. (2007). *Raising Student Achievement in Latin America: The Challenge for the 21st Century*. Washington, D.C.: Latin American Development Forum. Banco Mundial.
- Walker, M. (2011). *PISA 2009 Plus Results: Performance of 15-year old in reading, mathematics and science for 10 additional participants*. Australian Council for Educational Research.
- Zyphur, M., & Oswald, F. (2015). *Bayesian estimation and inference: A user's guide*. *Journal of Management*, 41: 390-420.

8 Anexos

Anexo 1 Distribuciones condicionales

Usando θ y W para denotar un parámetro desconocido y una matriz de covarianzas desconocida, respectivamente; y usando $p(\cdot)$, $E(\cdot)$ y $Var(\cdot)$ para denotar la función de densidad, la esperanza y la varianza, respectivamente:

1. Distribución Gamma:

$$\begin{aligned}\theta &\sim \text{Gamma}[\alpha, \beta] \\ p(\theta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{(\alpha-1)} e^{-\beta\theta} \\ E(\theta) &= \alpha/\beta \\ \text{Var}(\theta) &= \alpha/\beta^2\end{aligned}$$

2. Distribución Gamma invertida:

$$\begin{aligned}\theta &\sim \text{IG}[\alpha, \beta] \\ p(\theta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-(\alpha+1)} e^{-\beta/\theta} \\ E(\theta) &= \frac{\beta}{(\alpha-1)} \\ \text{Var}(\theta) &= \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}\end{aligned}$$

3. Relación entre la distribución Gamma y la distribución Gamma invertida:

$$\text{Si } \theta \sim \text{IG}[\alpha, \beta], \text{ entonces } \theta^{-1} \sim \text{Gamma}[\alpha, \beta]$$

4. Distribución Wishart:

$$\begin{aligned}W &\sim W_q[R_0, \rho_0] \\ p(W) &= \left[2^{\frac{\rho_0 q}{2}} \pi^{\frac{q(q-1)}{4}} \sum_{i=1}^q \Gamma\left(\frac{\rho_0 + 1 - i}{2}\right) \right]^{-1} \\ &\times |R_0|^{-\rho_0/2} \times |W|^{(\rho_0 - q - 1)} \times \exp\left\{-\frac{1}{2} \text{tr}(R_0^{-1}W)\right\} \\ E(W) &= \rho_0 R_0\end{aligned}$$

5. Distribución Wishart invertida:

$$W \sim \text{IW}_q[R_0^{-1}, \rho_0]$$

$$\begin{aligned}
 p(W) &= \left[2^{\frac{\rho_0 q}{2}} \pi^{\frac{q(q-1)}{4}} \sum_{i=1}^q \Gamma\left(\frac{\rho_0 + 1 - i}{2}\right) \right]^{-1} \\
 &\times |R_0|^{-\rho_0/2} \times |W|^{-(\rho_0 - q - 1)} \times \exp\left\{-\frac{1}{2} \text{tr}(R_0^{-1} W^{-1})\right\} \\
 E(W) &= \frac{R_0^{-1}}{\rho_0 - q - 1}
 \end{aligned}$$

6. Relación entre la distribución Wishart y la distribución Wishart invertida:

Si $W \sim IW[R_0^{-1}, \rho_0]$, entonces $W^{-1} \sim W[R_0, \rho_0]$

Anexo 2: Gráficos de dispersión del nivel dentro de grupos

Gráfico 4 **Cargas factoriales directas, escenario 1:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

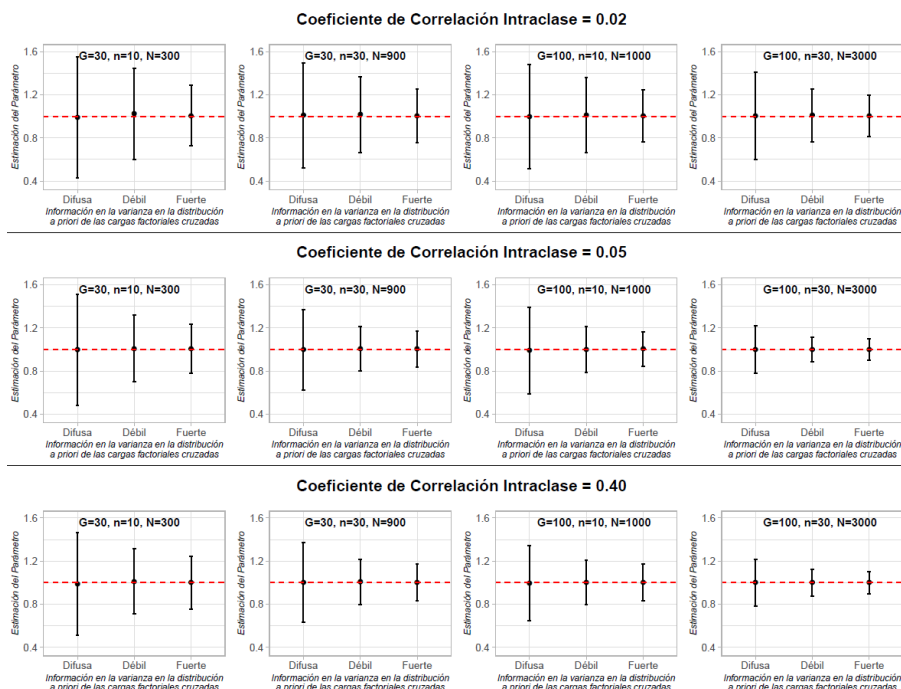


Gráfico 5 **Cargas factoriales directas, escenario 2:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

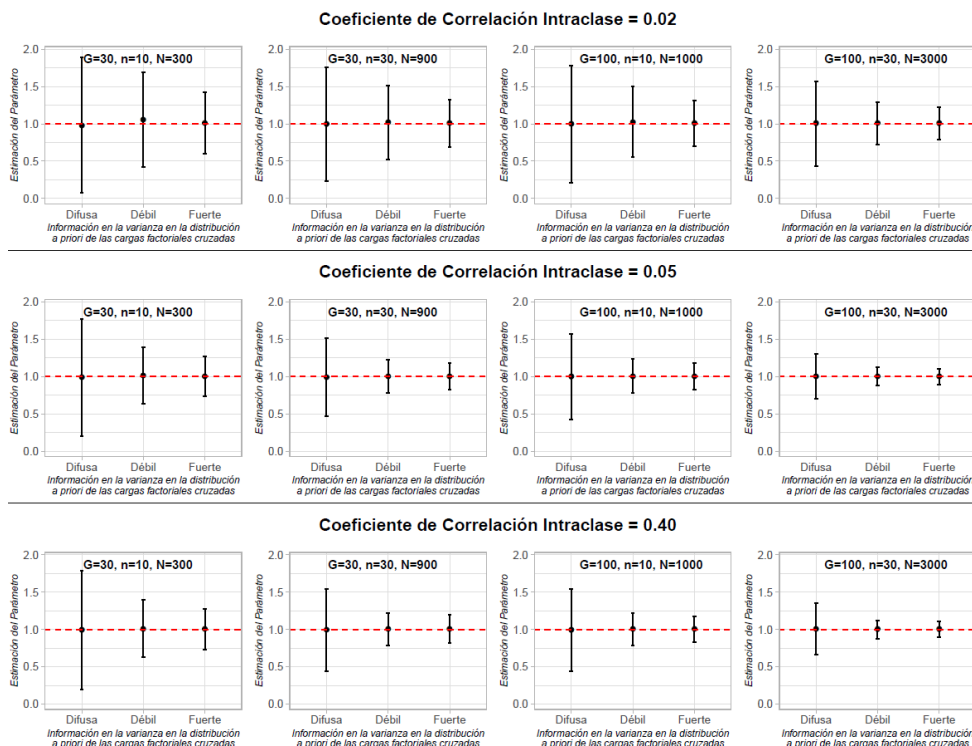


Gráfico 6 **Cargas factoriales directas, escenario 3:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

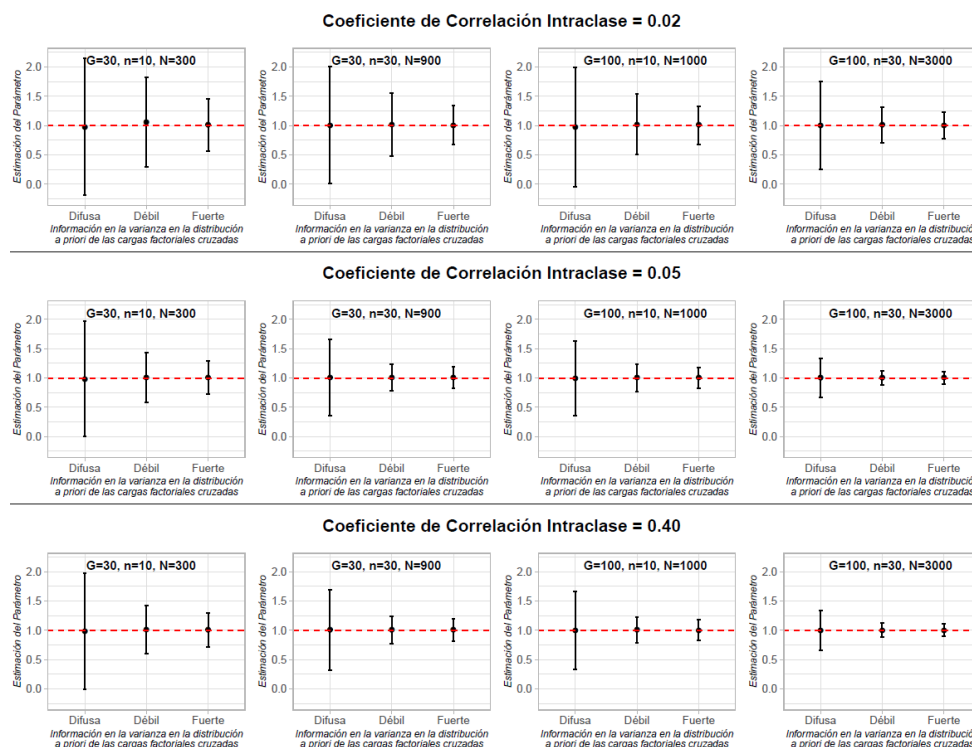


Gráfico 7 **Cargas factoriales directas, escenario 4:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

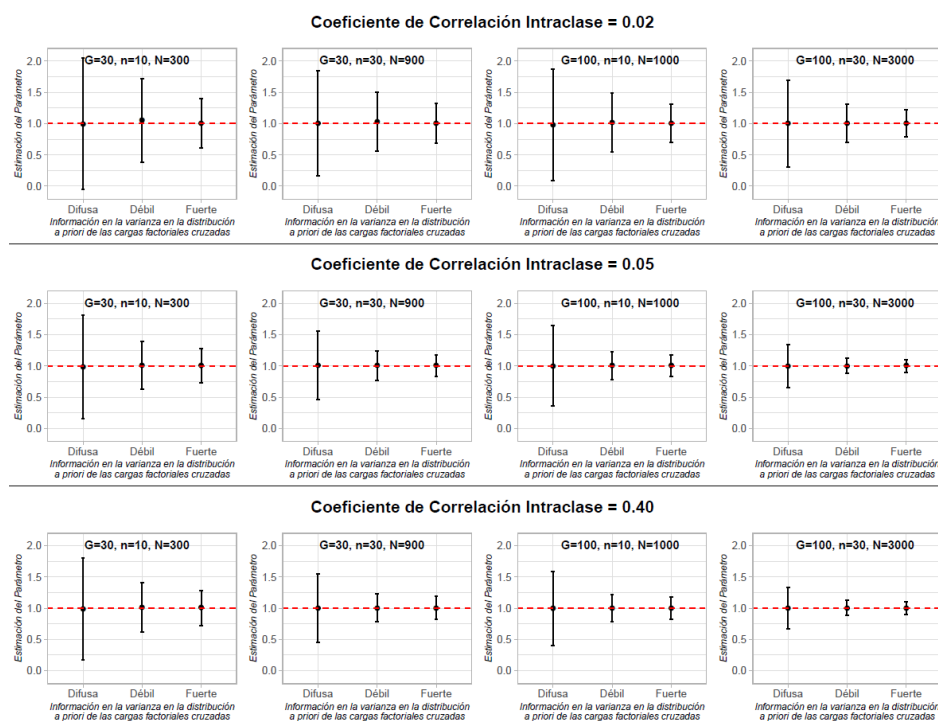


Gráfico 8 **Cargas factoriales directas, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

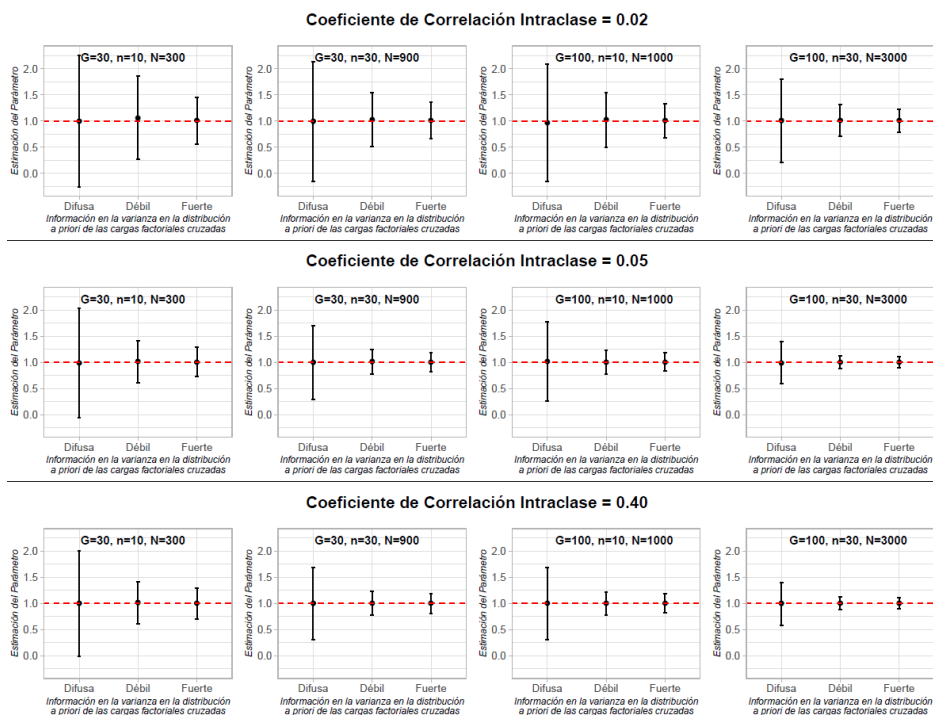


Gráfico 9 **Cargas factoriales directas, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

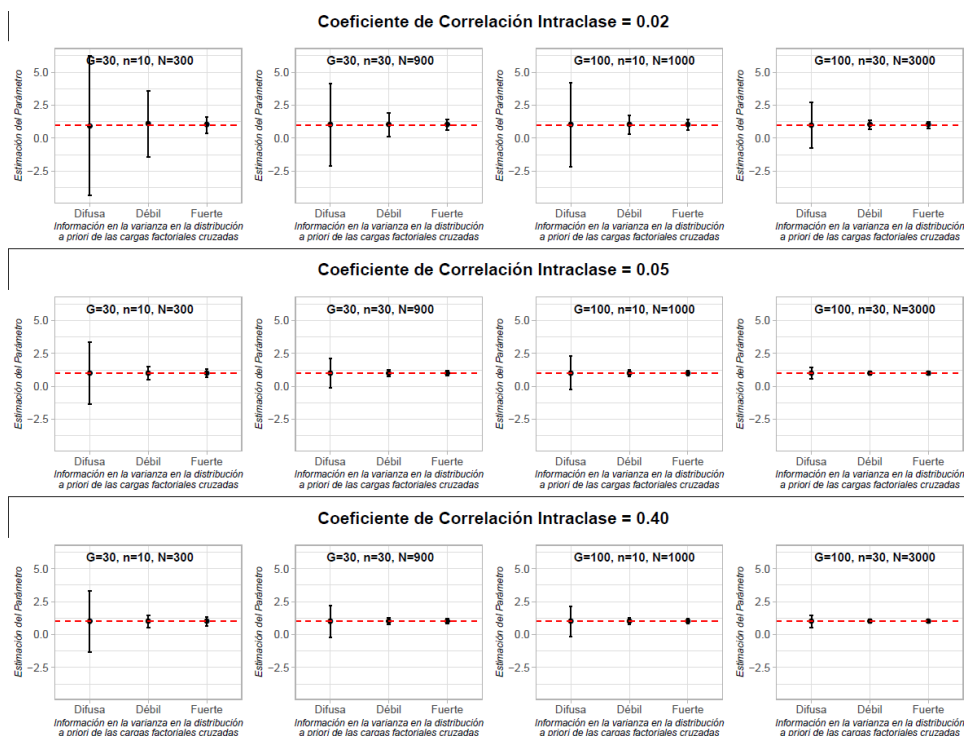


Gráfico 10 **Cargas factoriales cruzadas, escenario 1:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

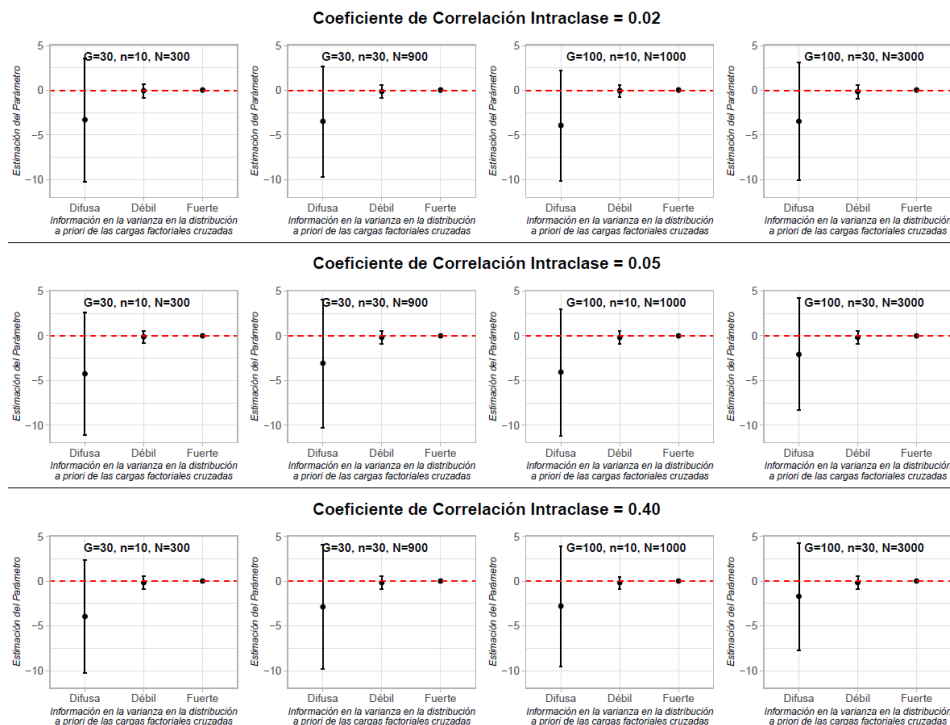


Gráfico 11 **Cargas factoriales cruzadas, escenario 2:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

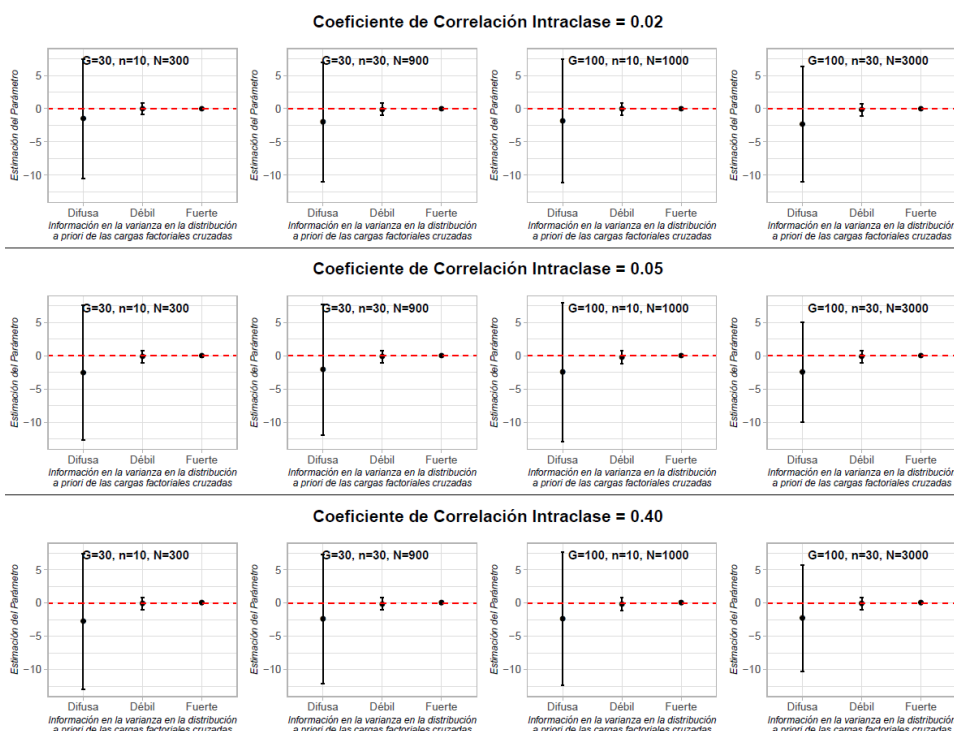


Gráfico 12 **Cargas factoriales cruzadas, escenario 3:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)

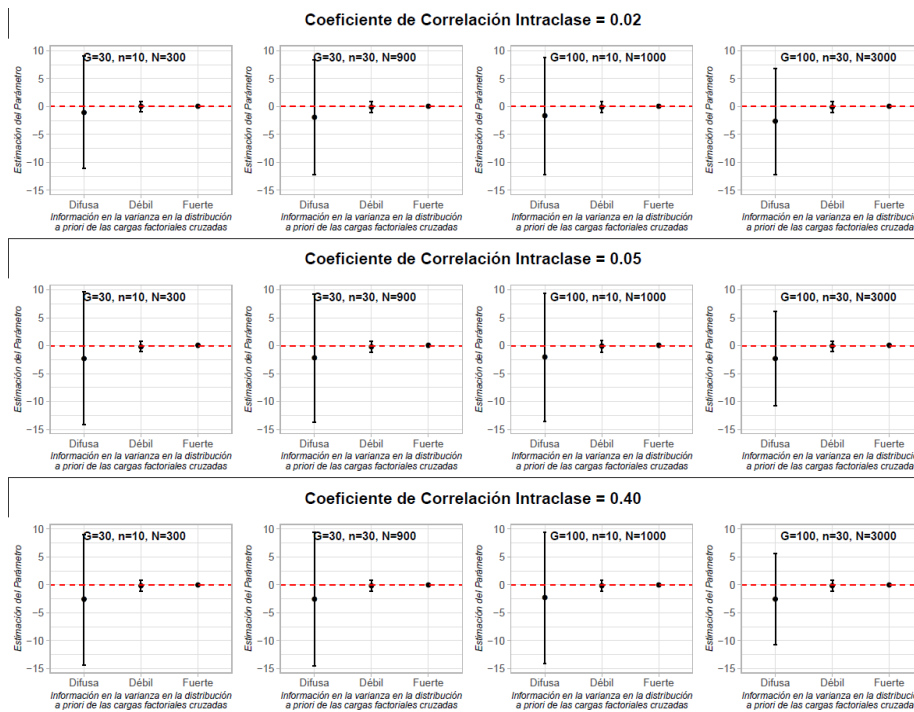


Gráfico 13 **Cargas factoriales cruzadas, escenario 4:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)

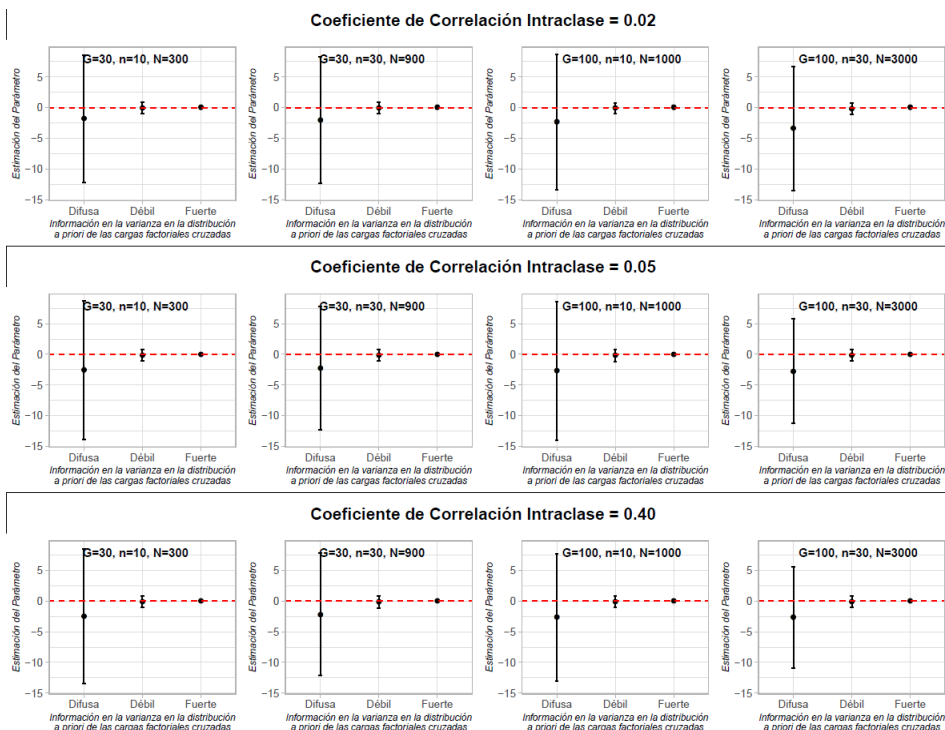


Gráfico 14 **Cargas factoriales cruzadas, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

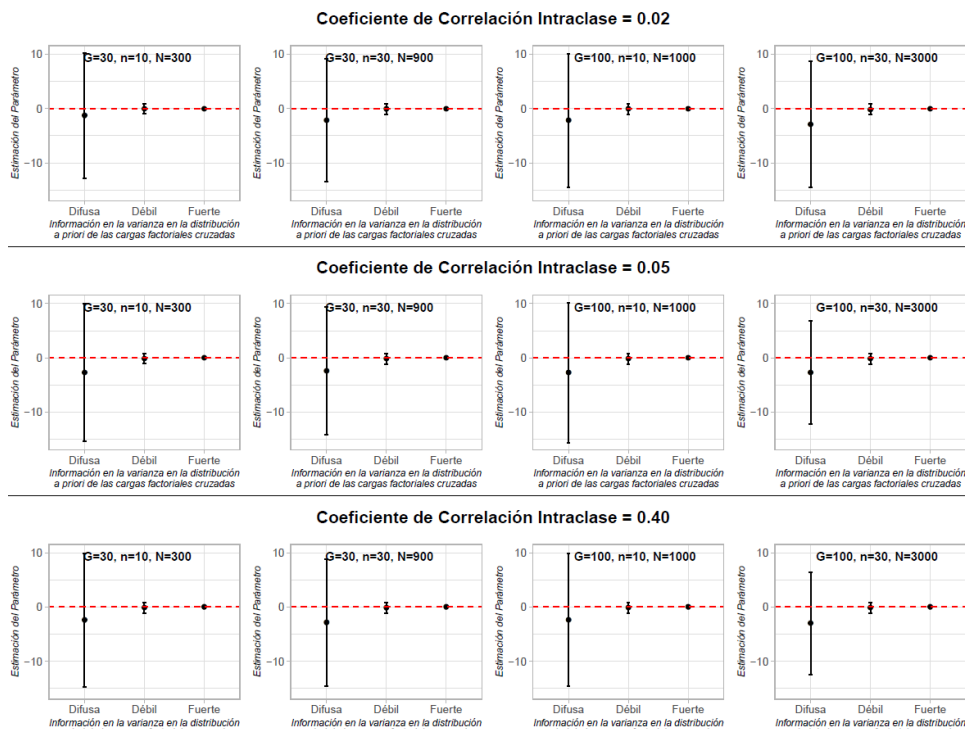


Gráfico 15 **Cargas factoriales cruzadas, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

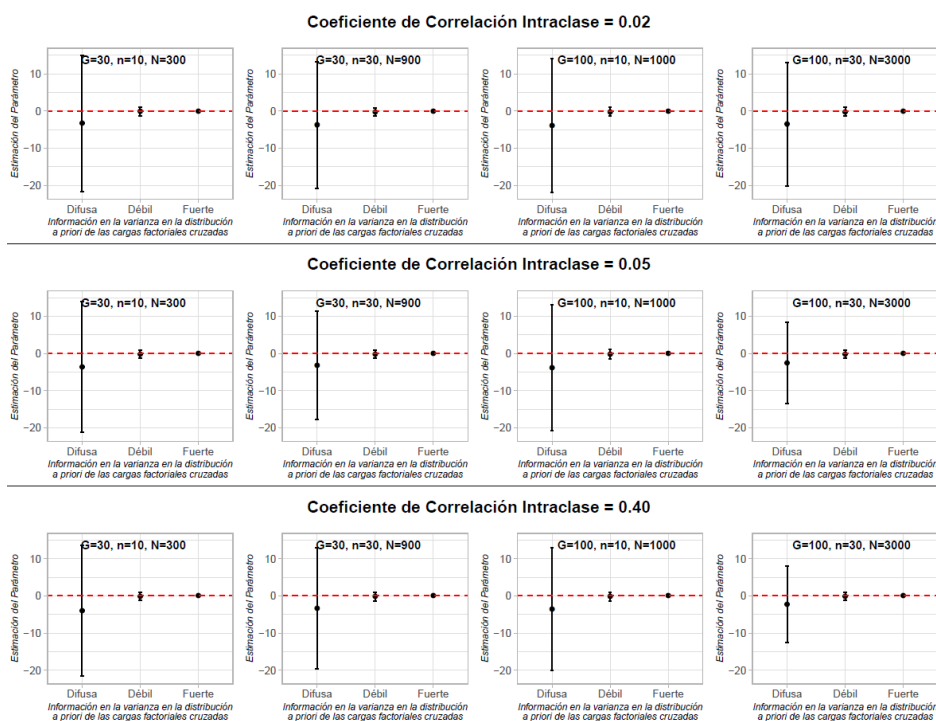


Gráfico 16 **Parámetro estructural, escenario 1:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

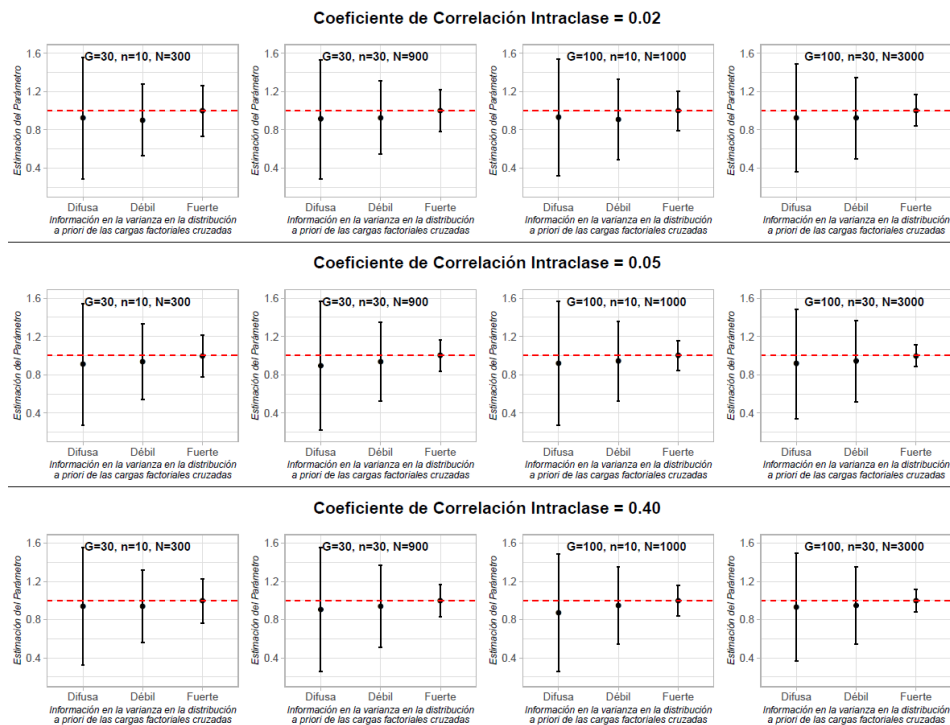


Gráfico 17 **Parámetro estructural, escenario 2:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

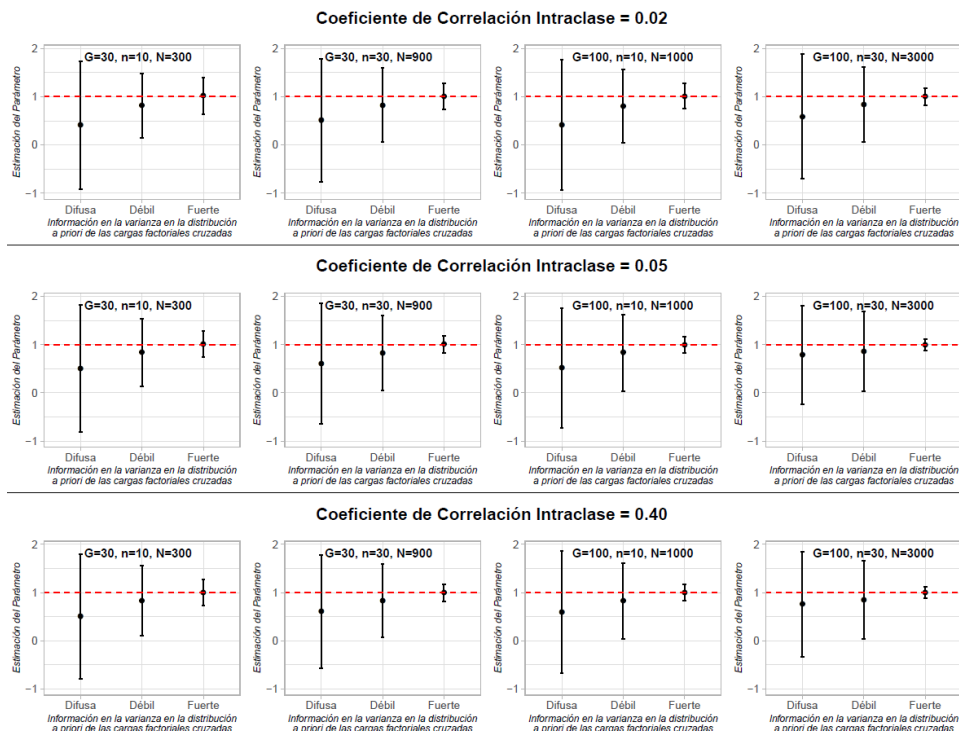


Gráfico 18 **Parámetro estructural, escenario 3:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

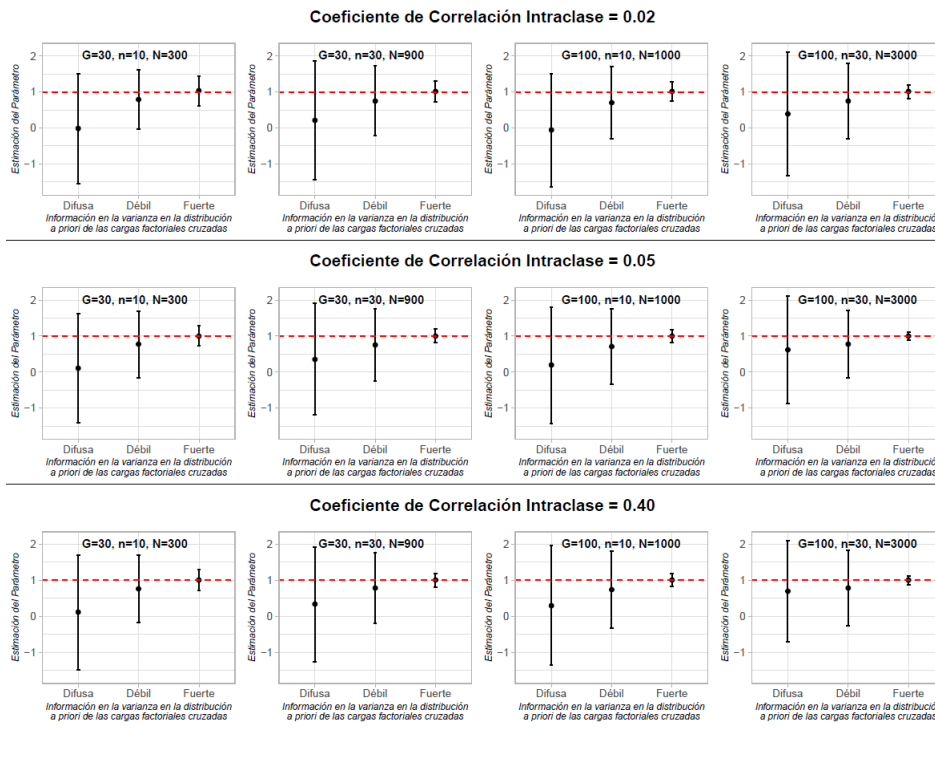


Gráfico 19 **Parámetro estructural, escenario 4:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

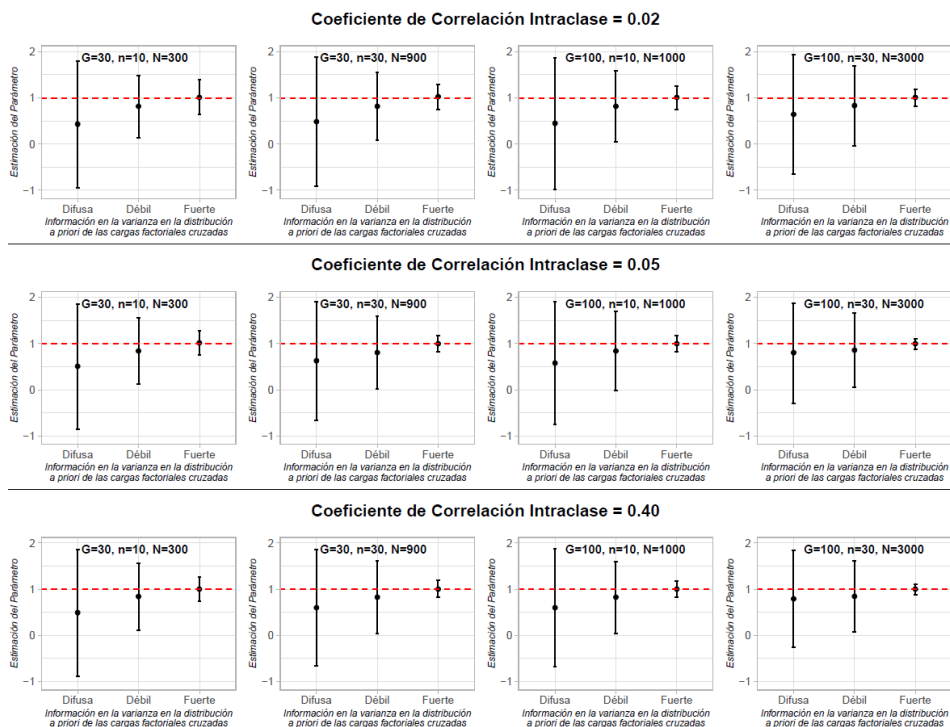


Gráfico 20 **Parámetro estructural, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)

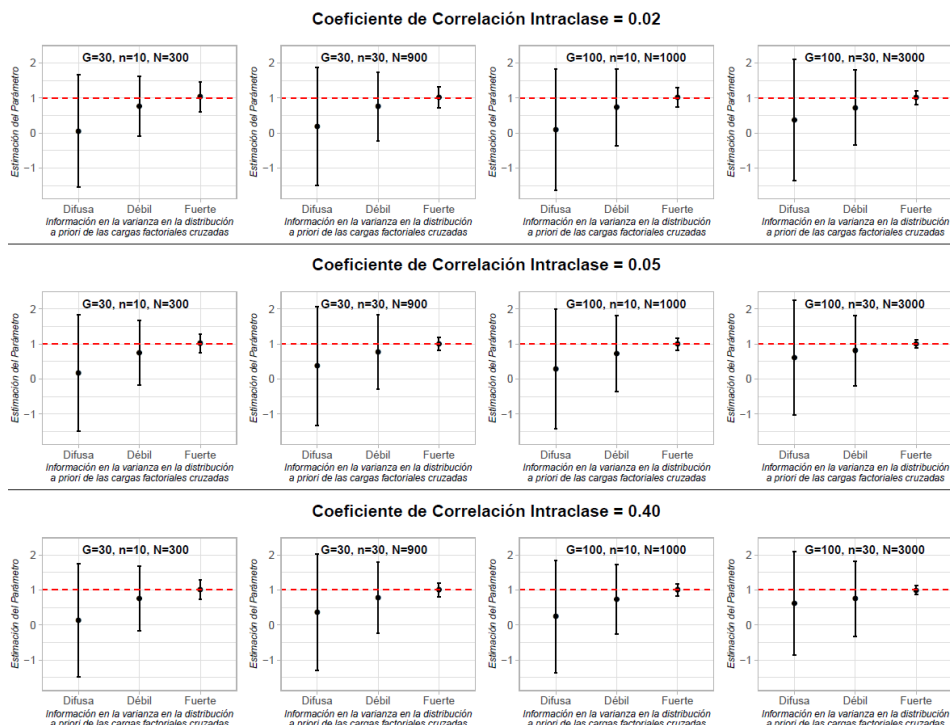


Gráfico 21 **Parámetro estructural, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)

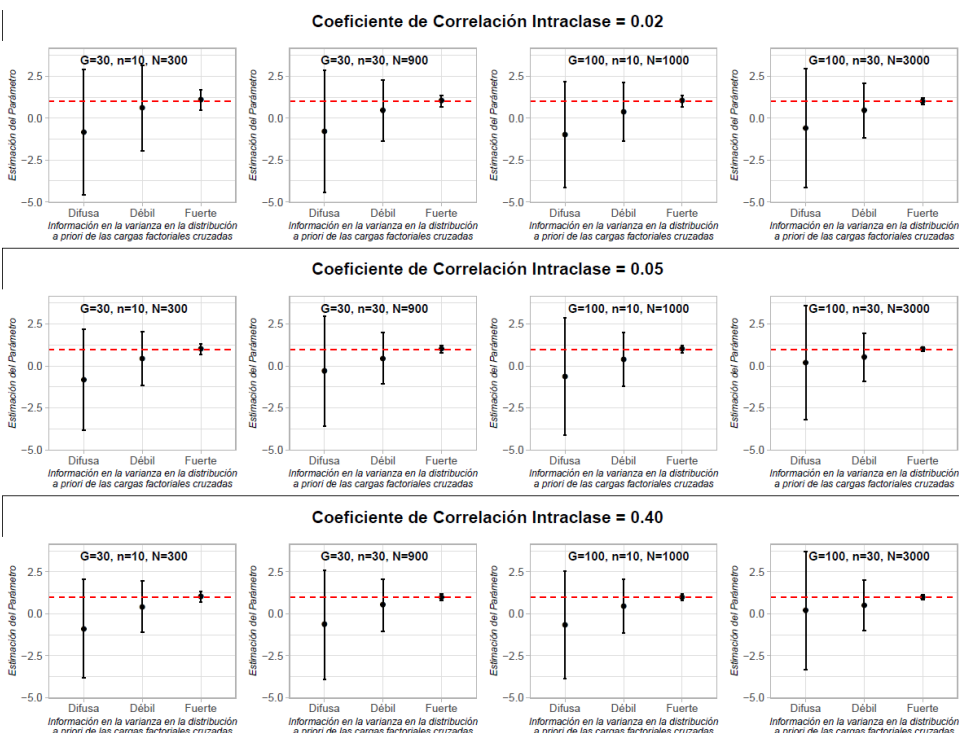


Gráfico 22 **Varianzas residuales, escenario 1:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

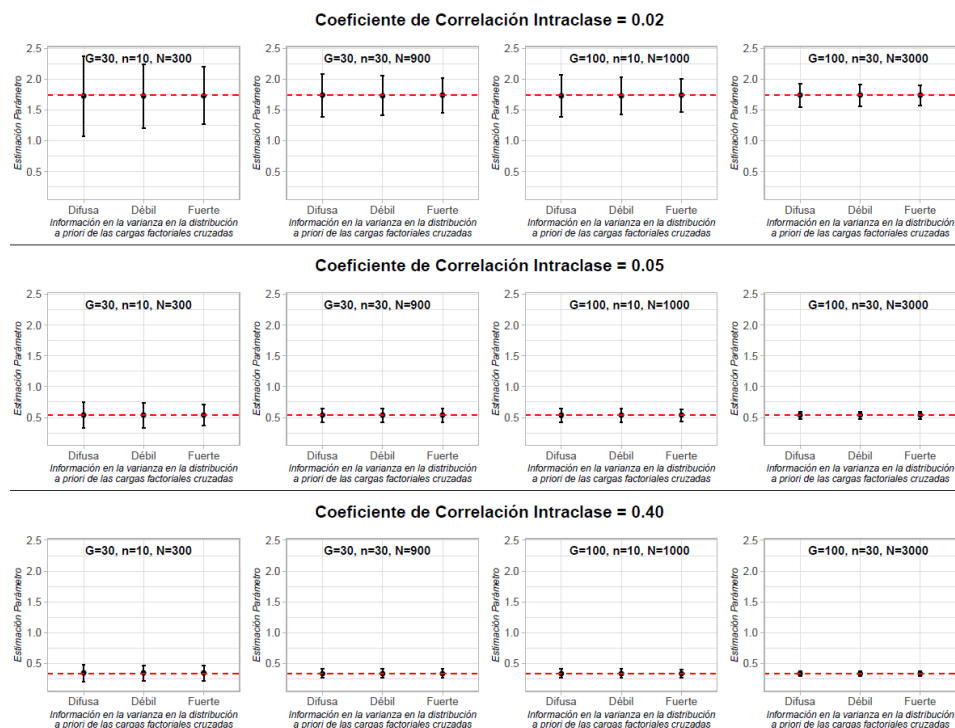


Gráfico 23 **Varianzas residuales, escenario 2:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

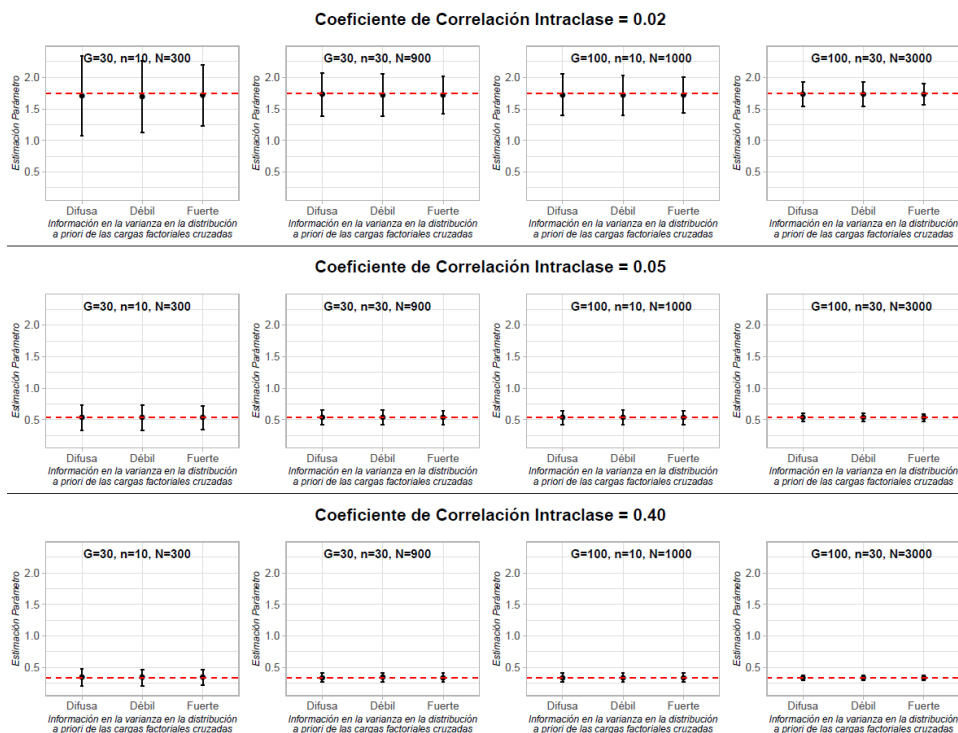


Gráfico 24 **Varianzas residuales, escenario 3:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

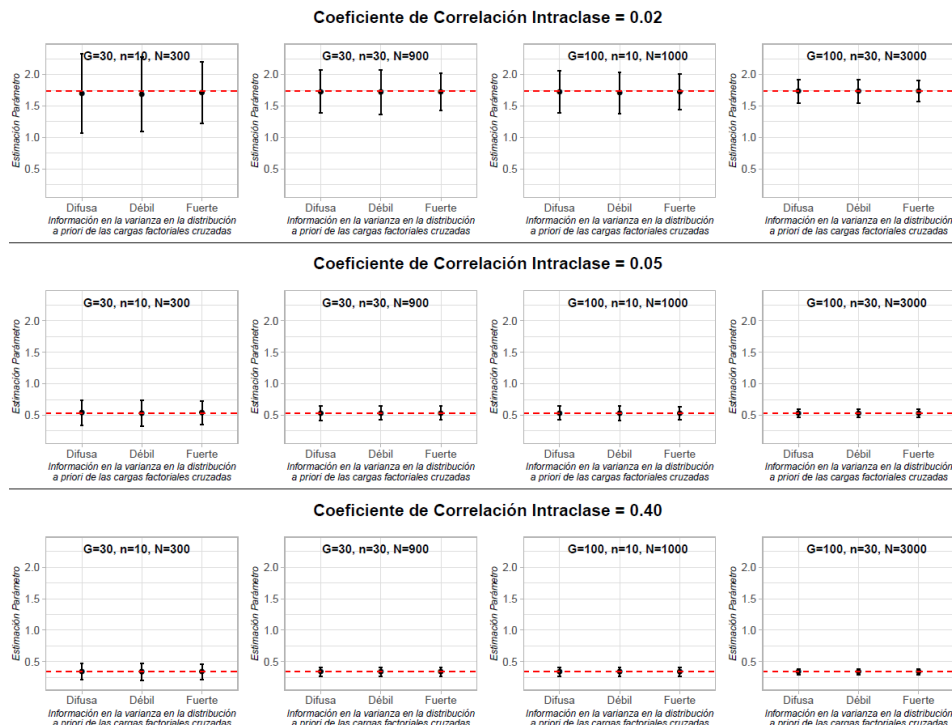


Gráfico 25 **Varianzas residuales, escenario 4:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

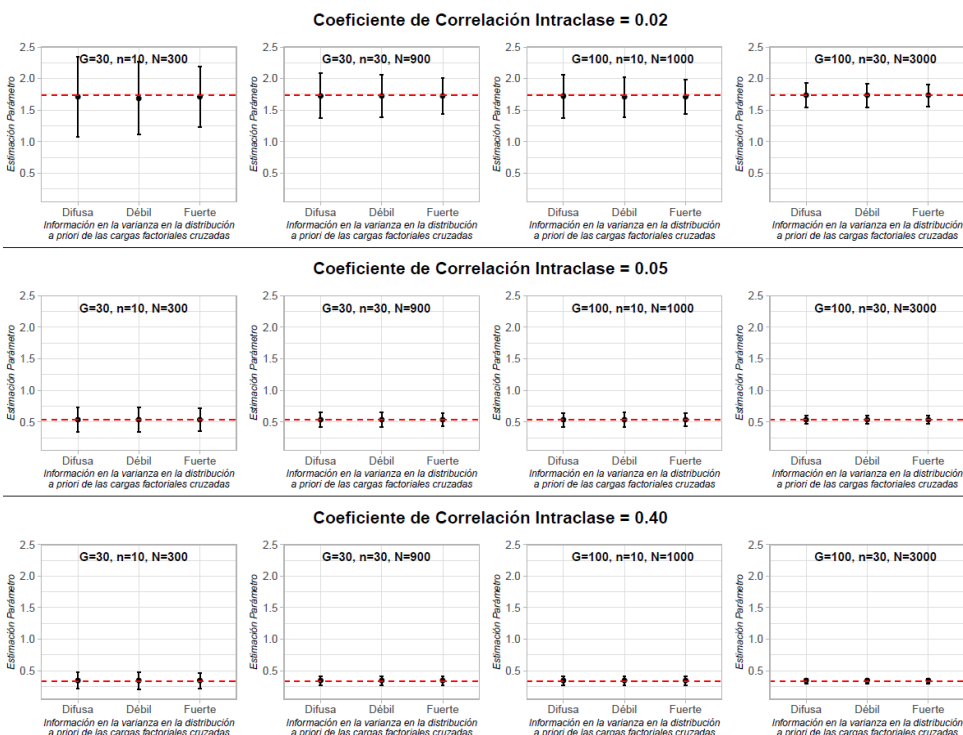


Gráfico 26 **Varianzas residuales, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)

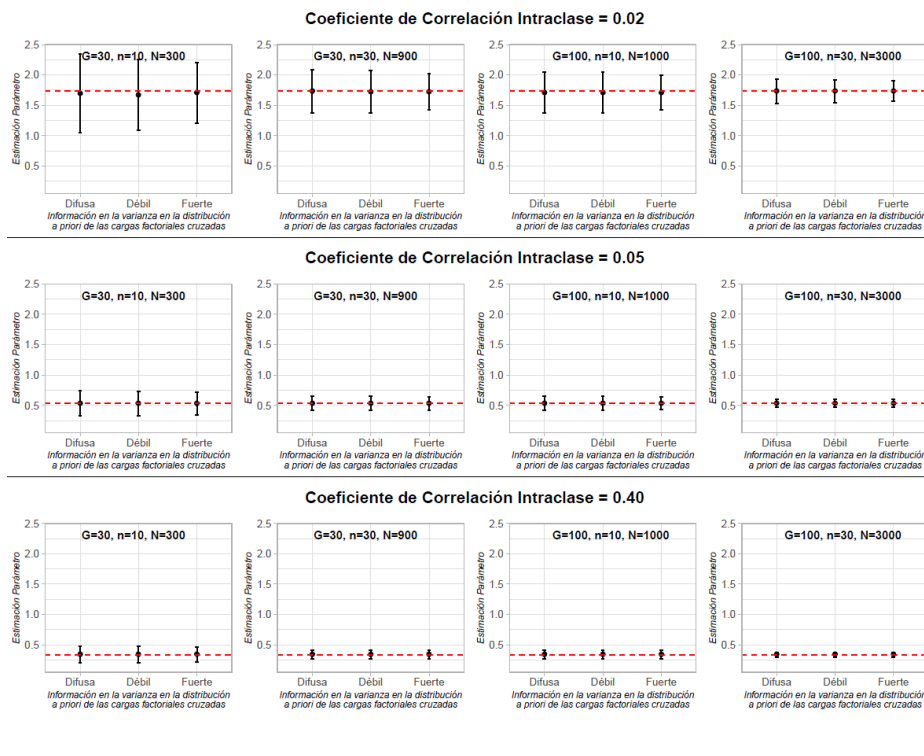


Gráfico 27 **Varianzas residuales, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)

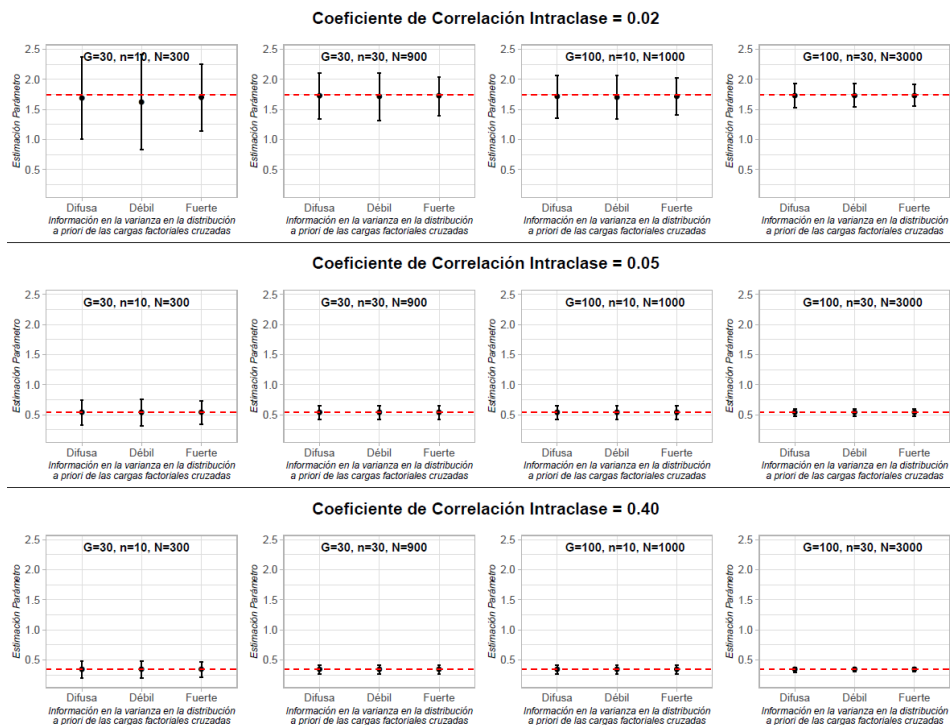


Gráfico 28 **Varianza factorial, escenario 1:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

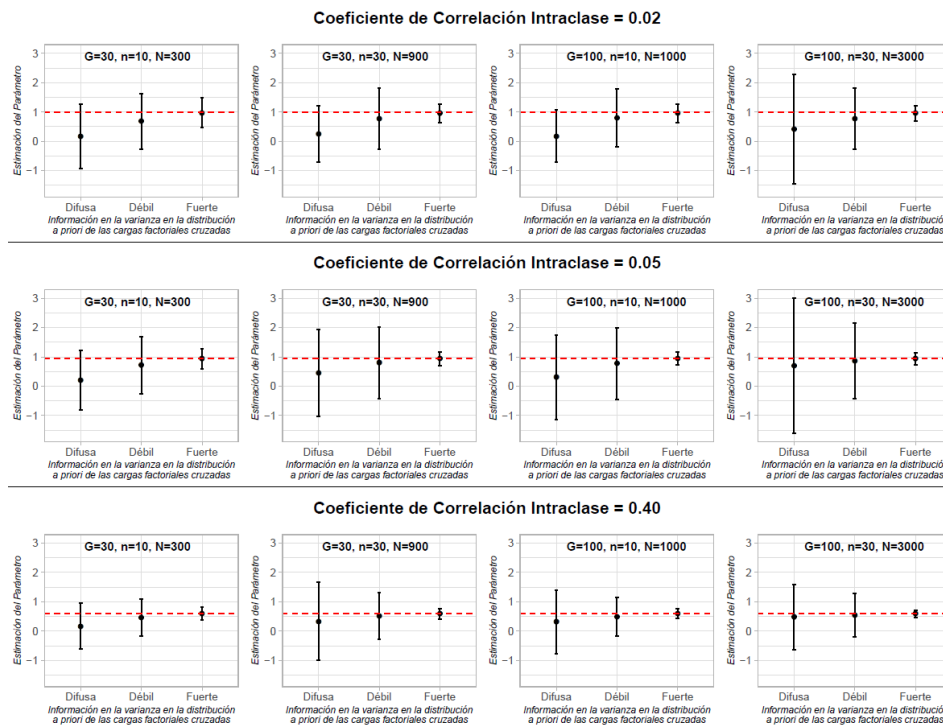


Gráfico 29 **Varianza factorial, escenario 2:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

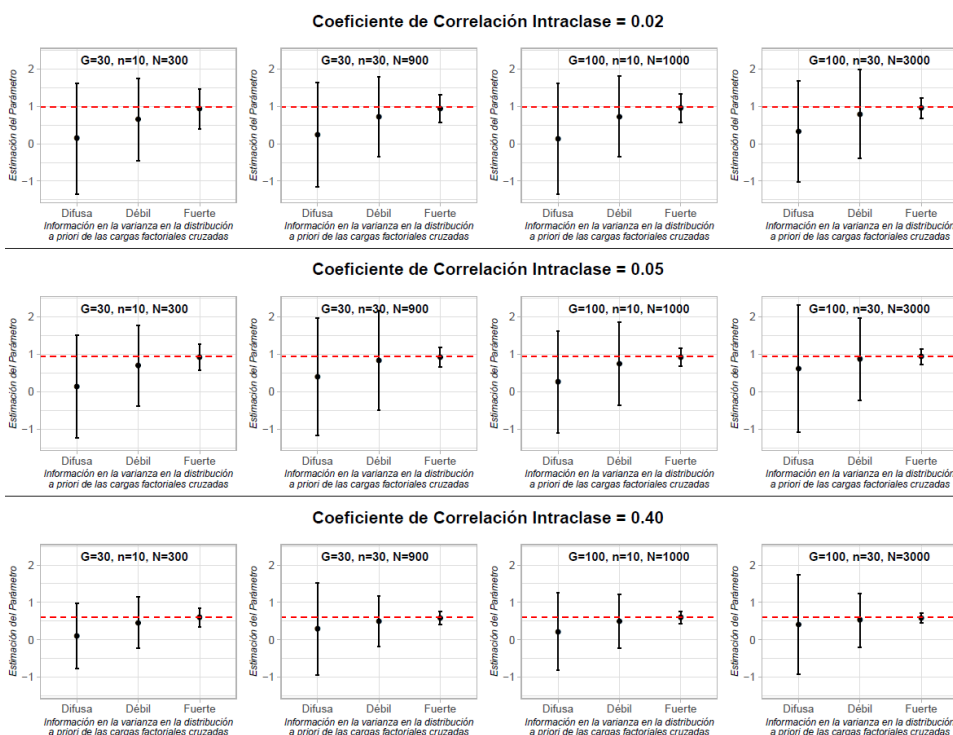


Gráfico 30 **Varianza factorial, escenario 3:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

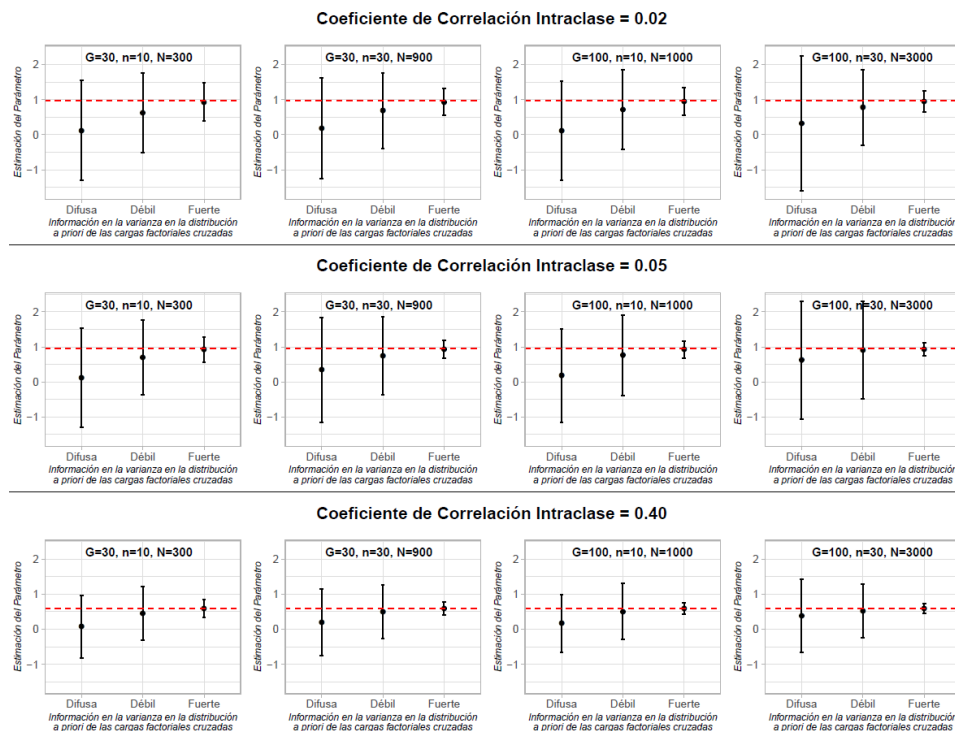


Gráfico 31 **Varianza factorial, escenario 4:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel dentro de grupos)

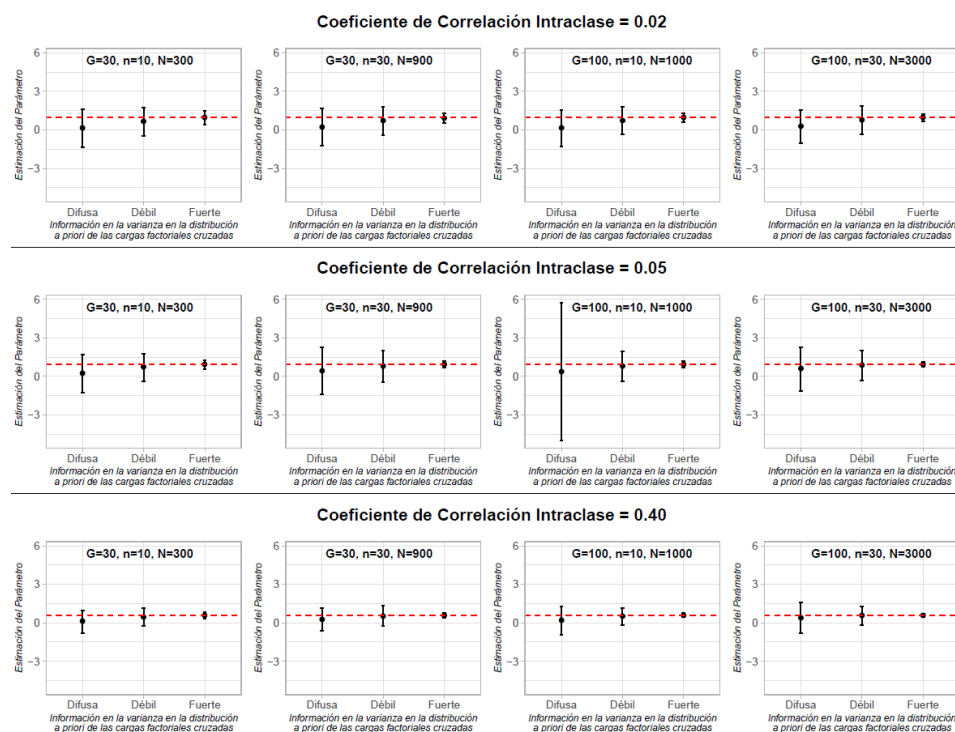


Gráfico 32 **Varianza factorial, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)

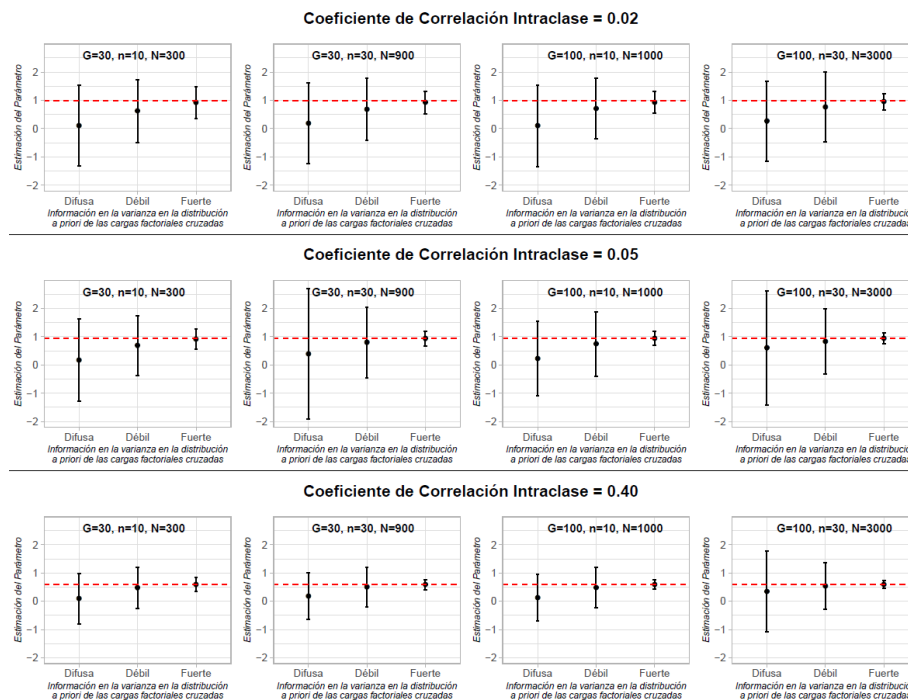
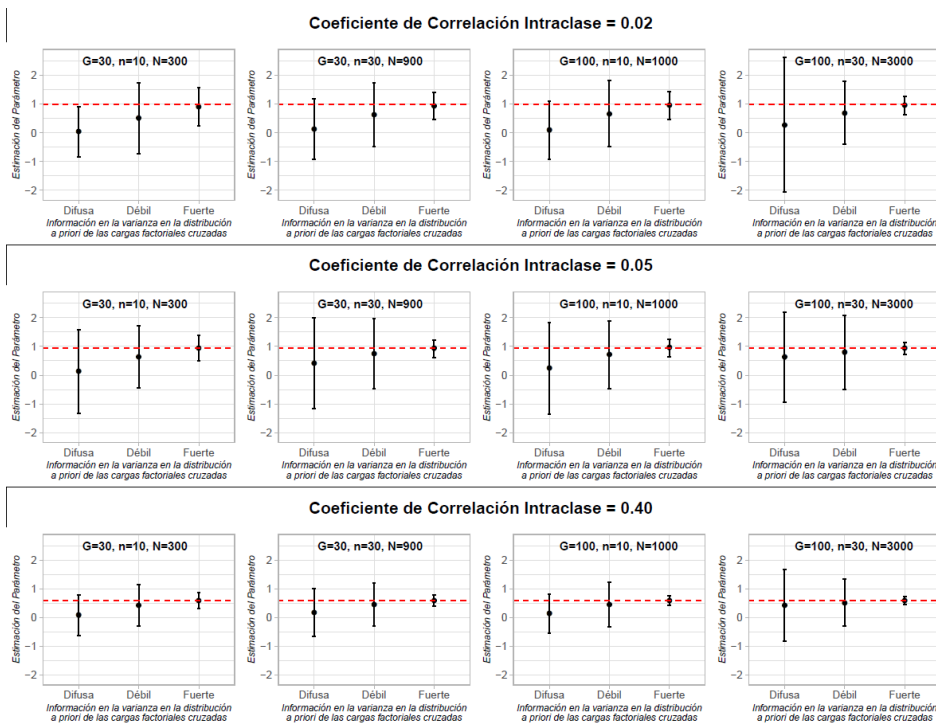


Gráfico 33 **Varianza factorial, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel dentro de grupos)



Anexo 3: Gráficos de dispersión del nivel entre grupos

Gráfico 34 *Cargas factoriales directas, escenario 1: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)*

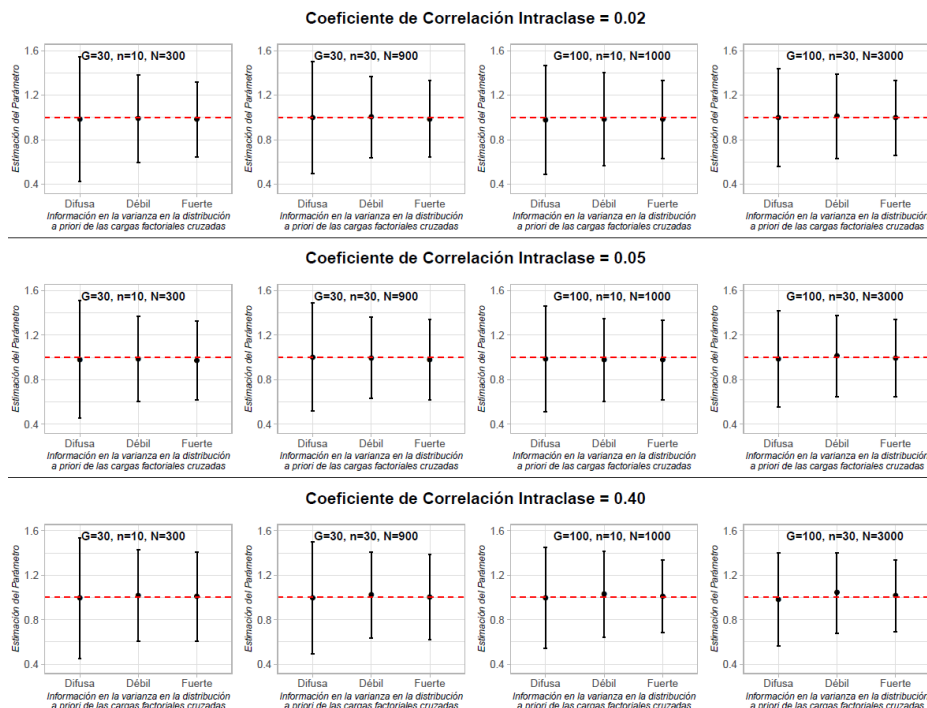


Gráfico 35 *Cargas factoriales directas, escenario 2: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)*

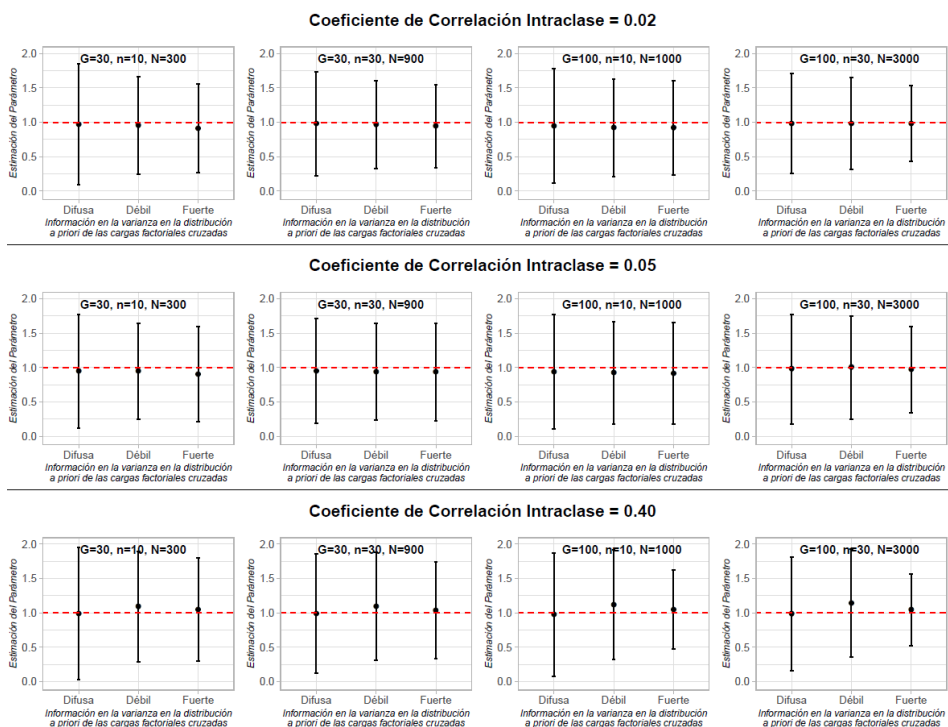


Gráfico 36 *Cargas factoriales directas, escenario 3: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)*

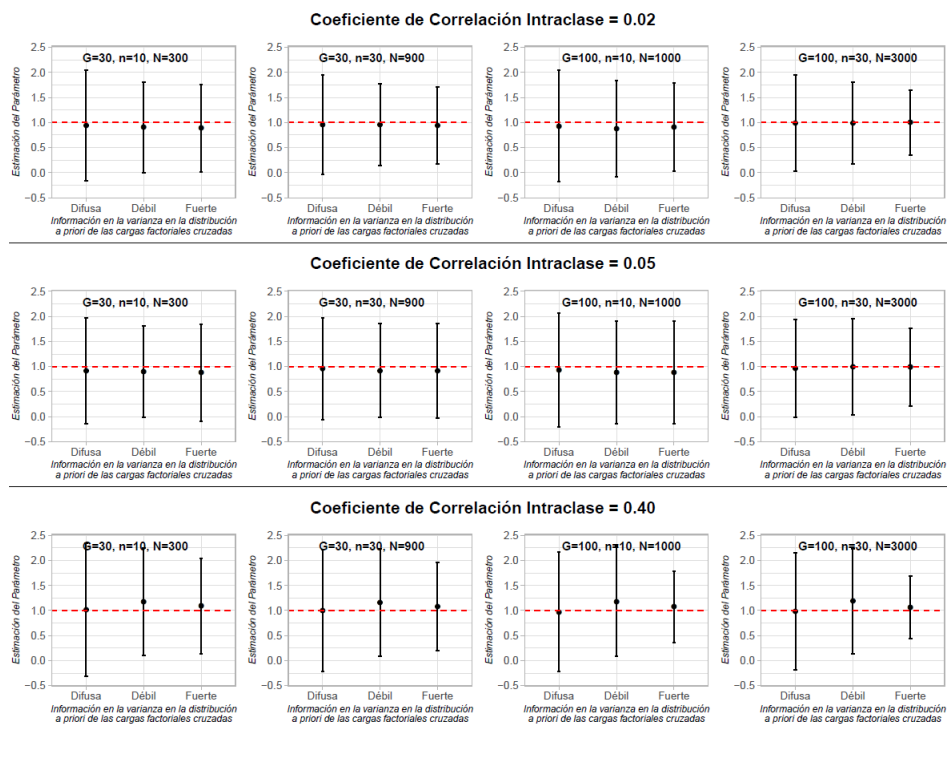


Gráfico 37 *Cargas factoriales directas, escenario 4: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)*

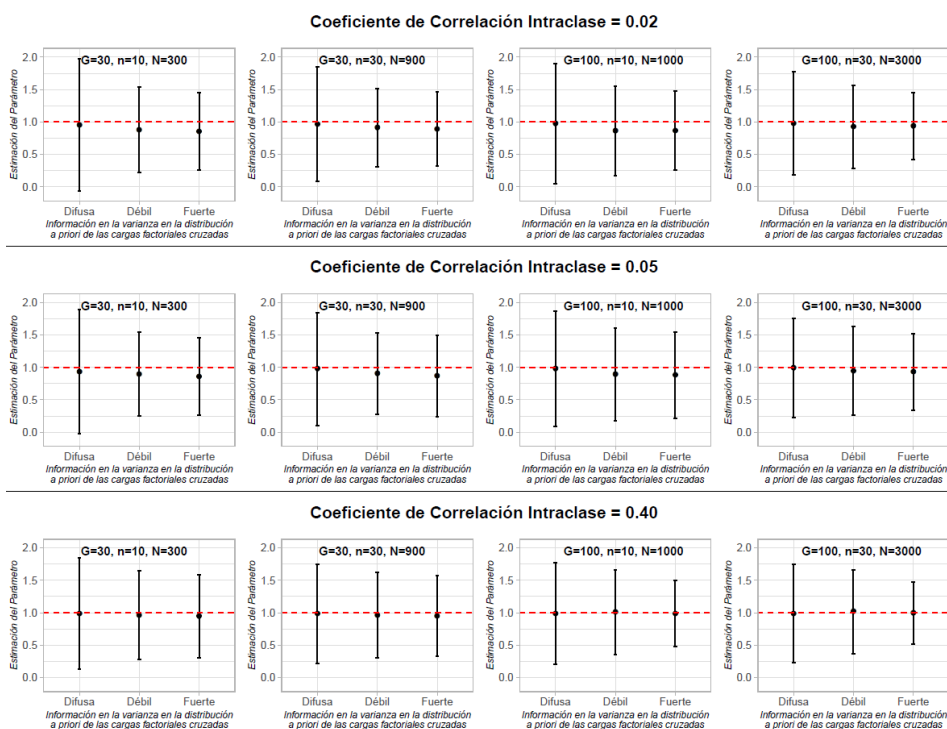


Gráfico 38 **Cargas factoriales directas, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

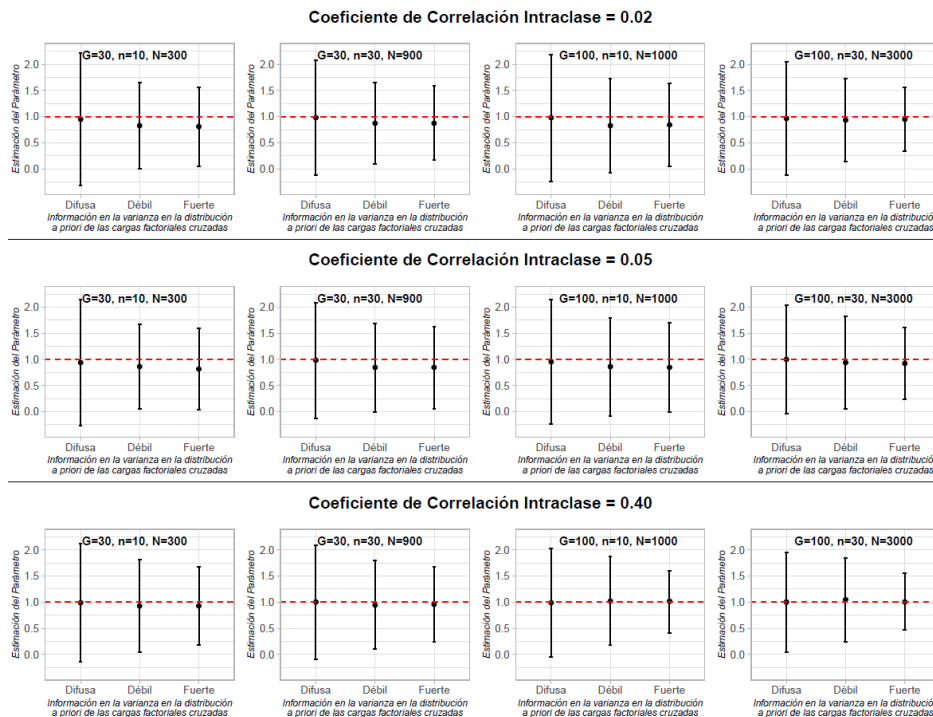


Gráfico 39 **Cargas factoriales directas, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

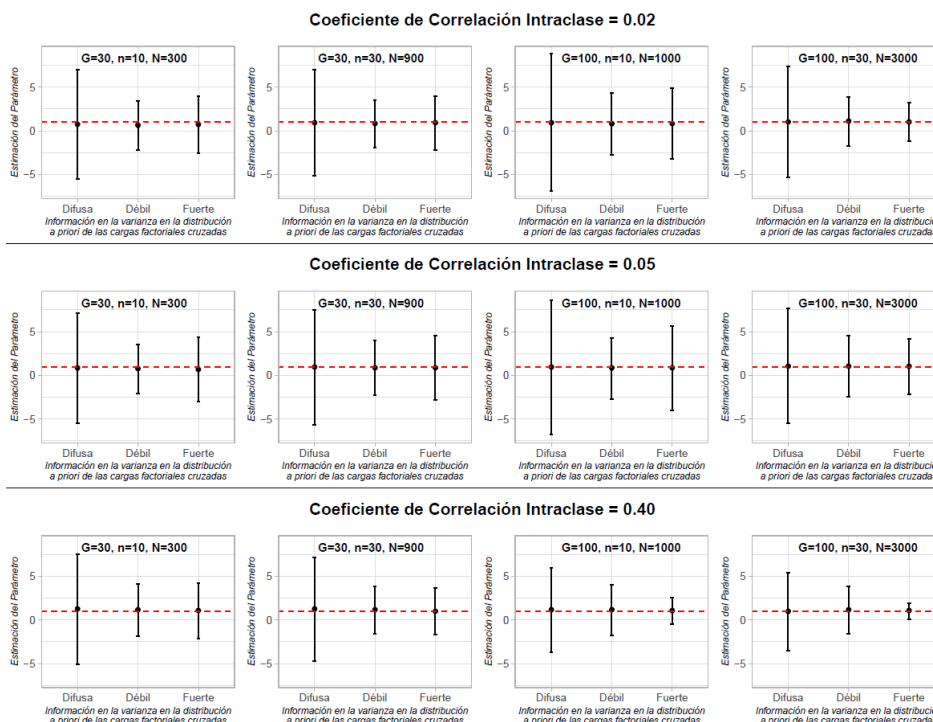


Gráfico 40 **Cargas factoriales cruzadas, escenario 1:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

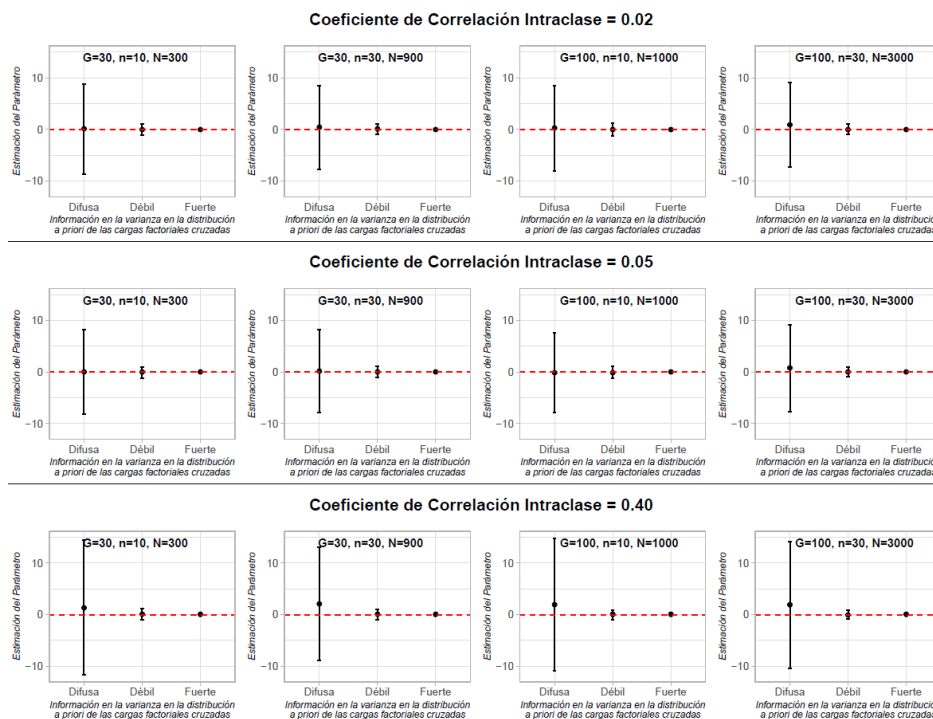


Gráfico 41 **Cargas factoriales cruzadas, escenario 2:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

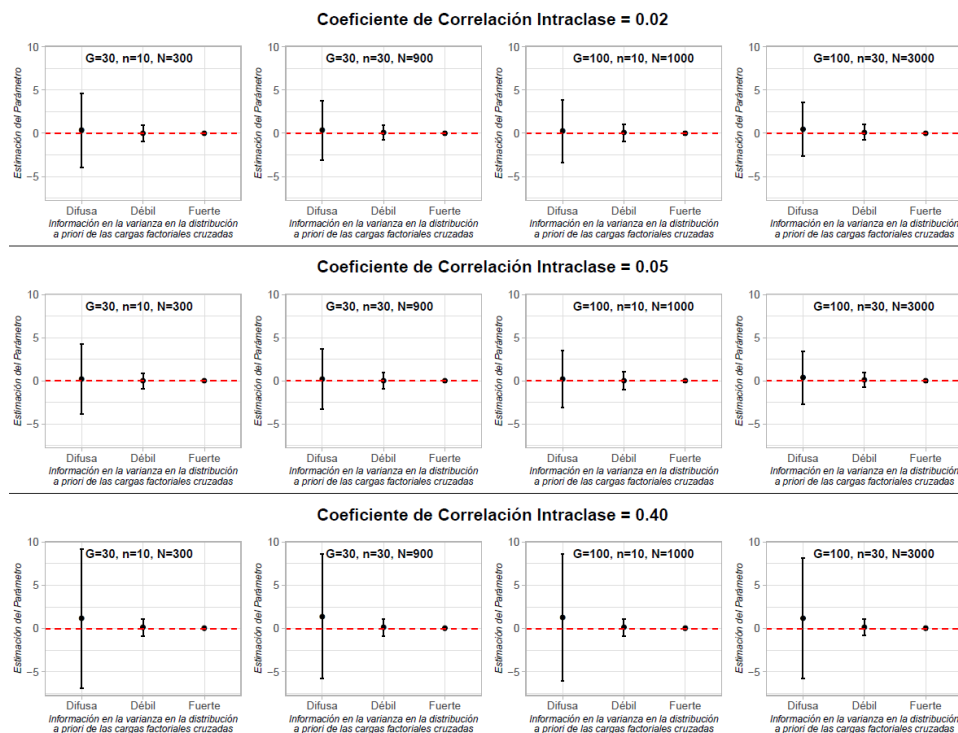


Gráfico 42 **Cargas factoriales cruzadas, escenario 3:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

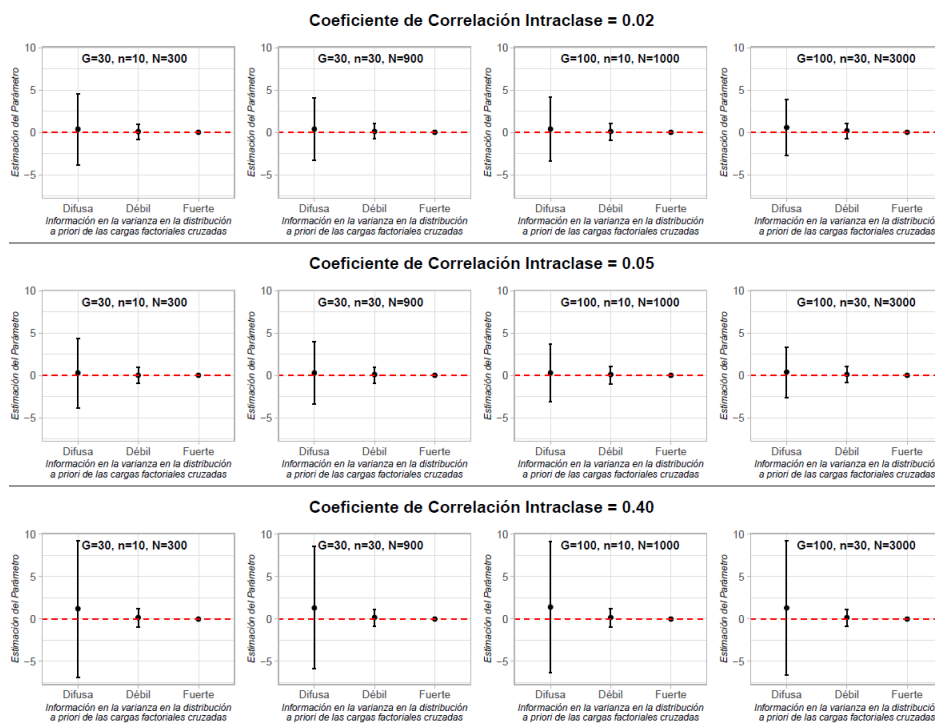


Gráfico 43 **Cargas factoriales cruzadas, escenario 4:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

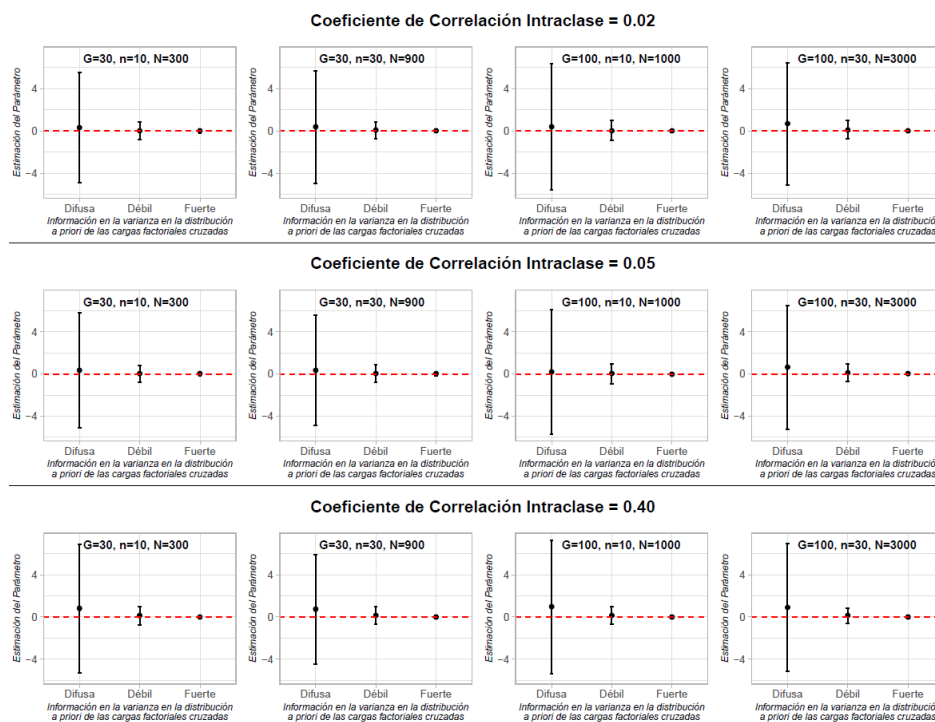


Gráfico 44 **Cargas factoriales cruzadas, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

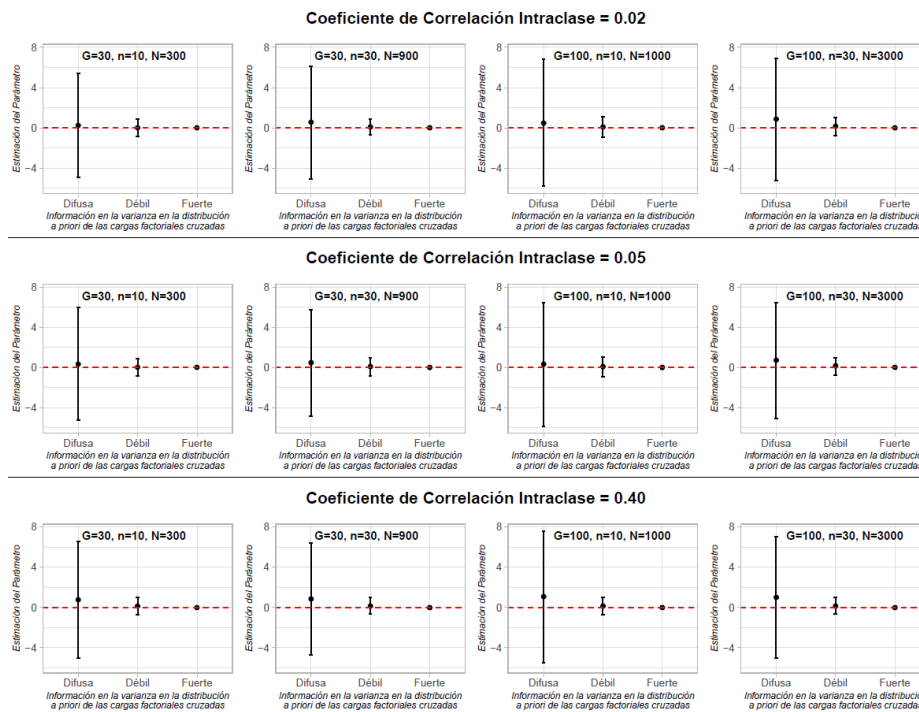


Gráfico 45 **Cargas factoriales cruzadas, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

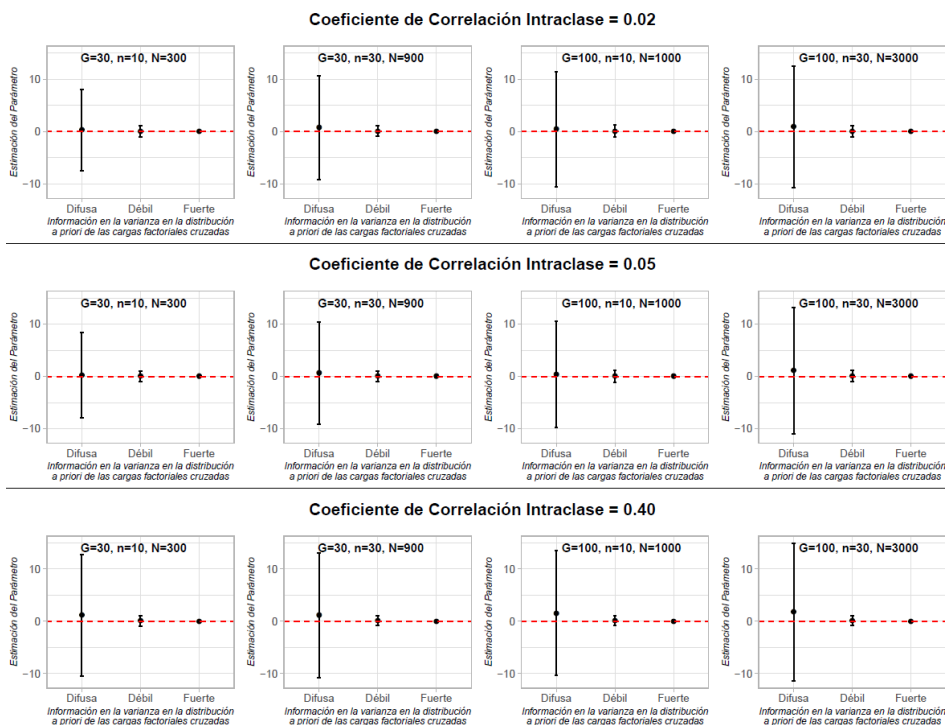


Gráfico 46 **Parámetro estructural, escenario 1:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

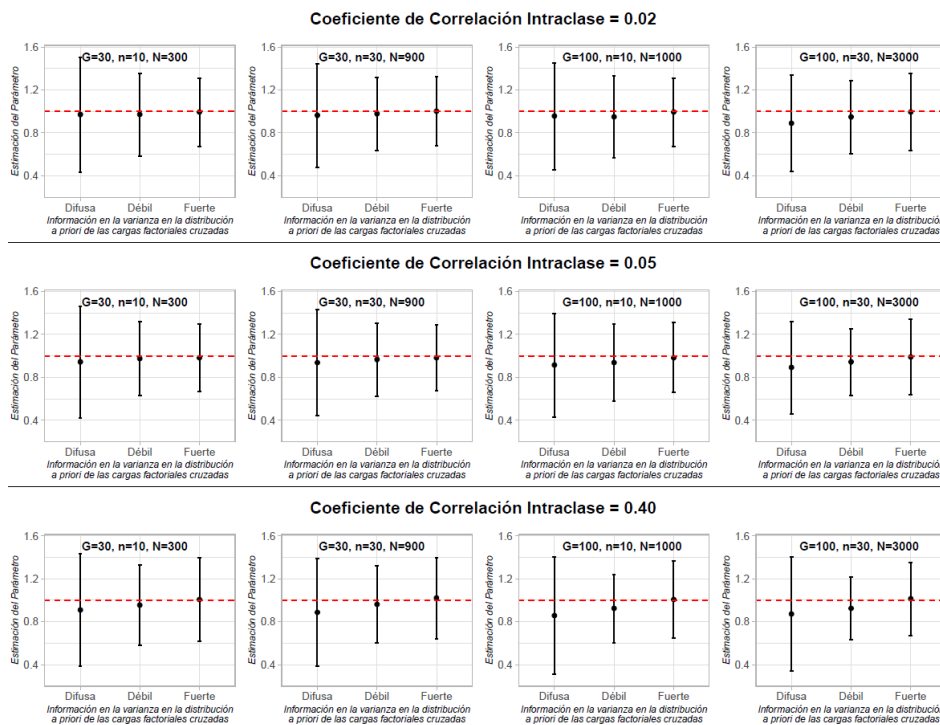


Gráfico 47 **Parámetro estructural, escenario 2:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

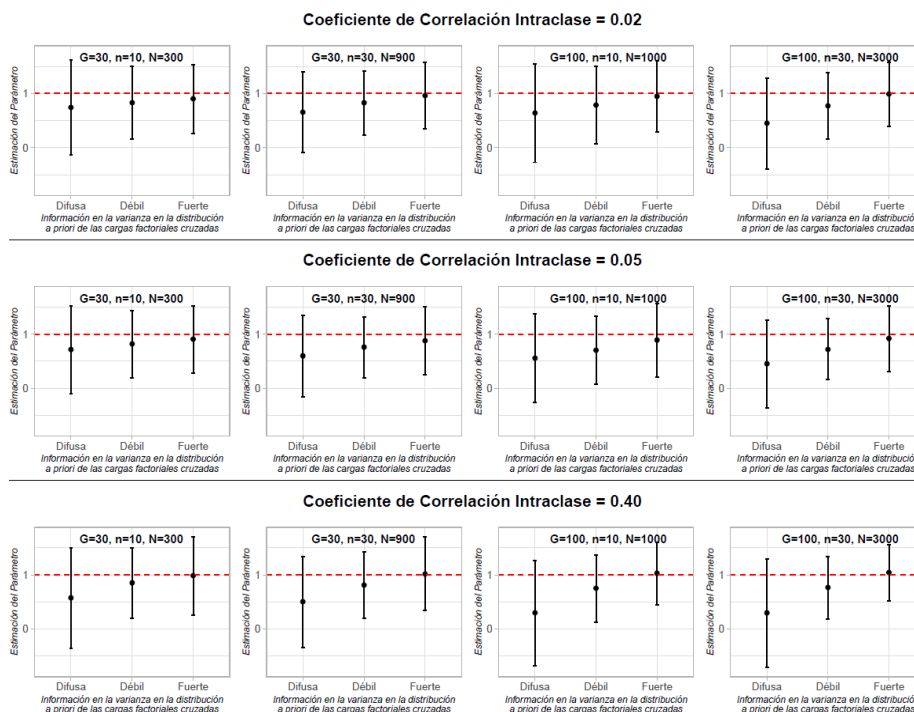


Gráfico 48 **Parámetro estructural, escenario 3:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

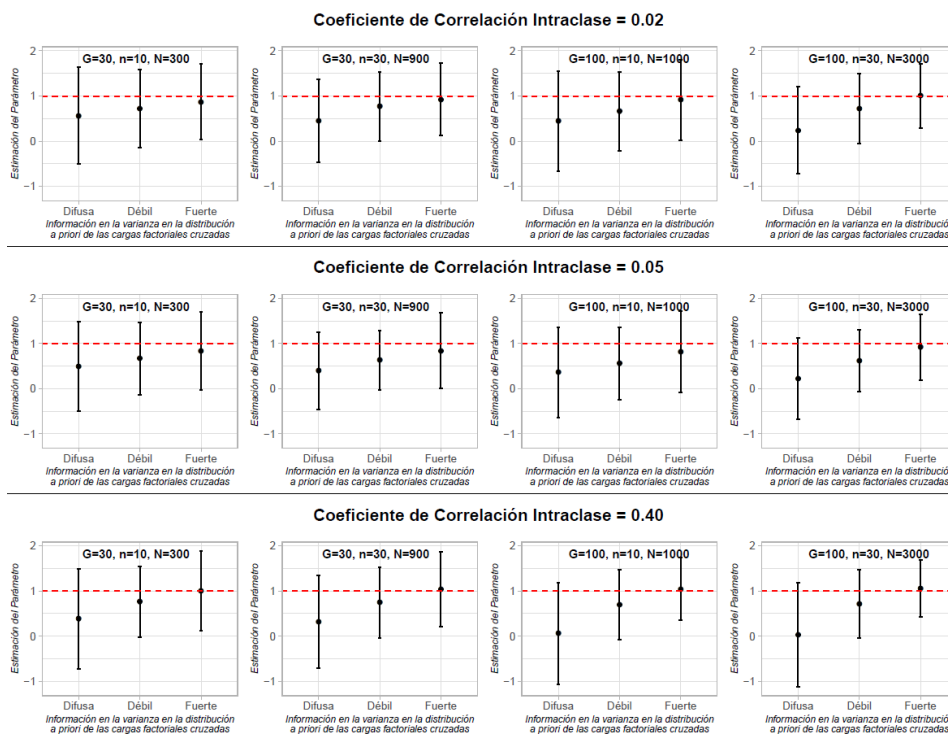


Gráfico 49 **Parámetro estructural, escenario 4:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

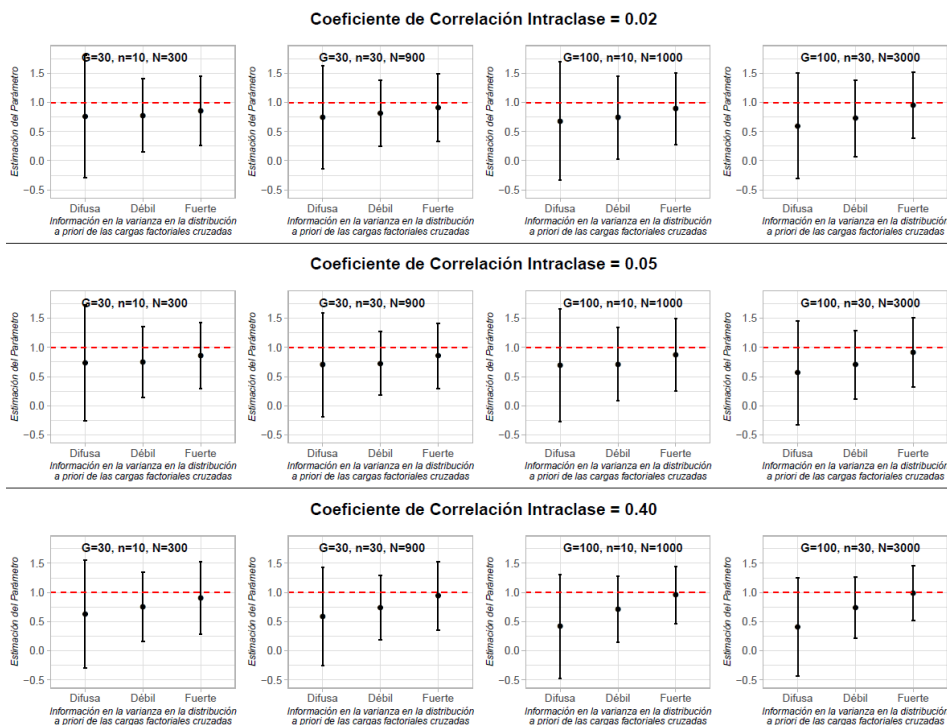


Gráfico 50 **Parámetro estructural, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

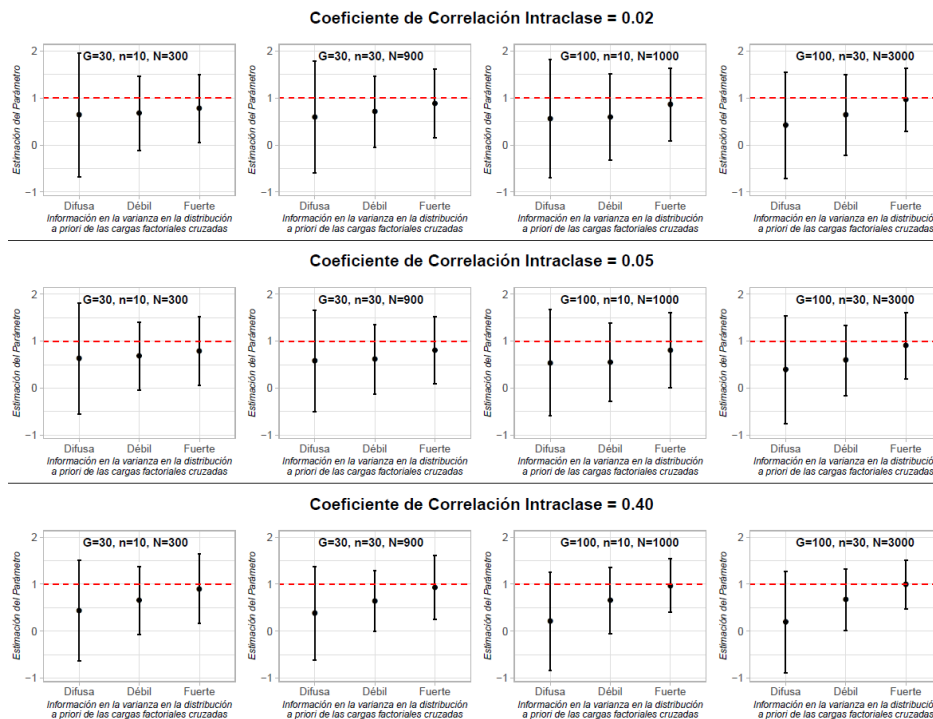


Gráfico 51 **Parámetro estructural, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

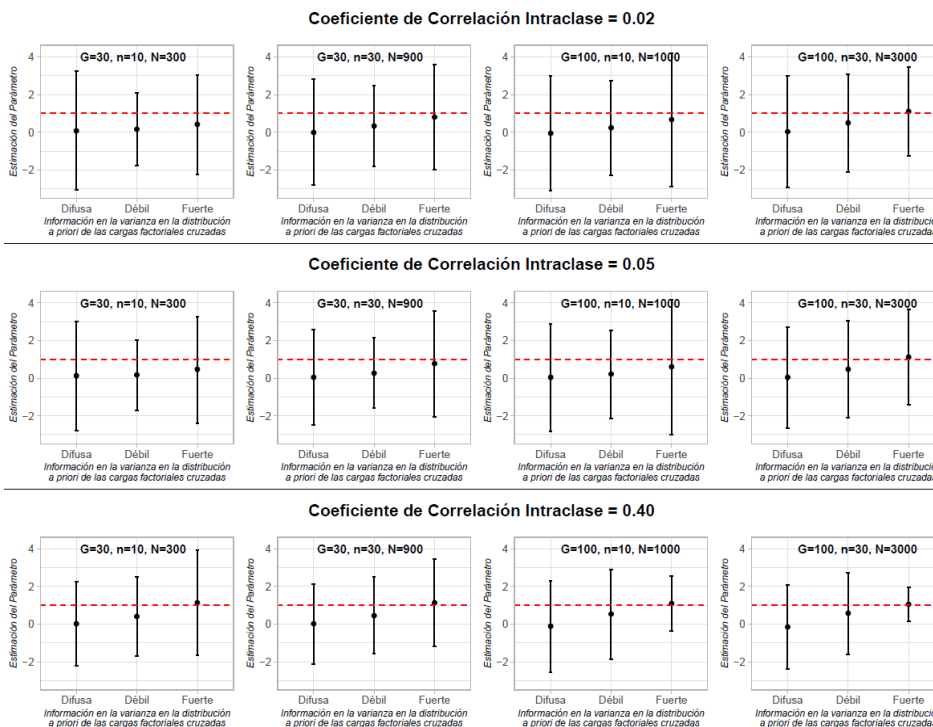


Gráfico 52 **Varianzas residuales, escenario 1:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

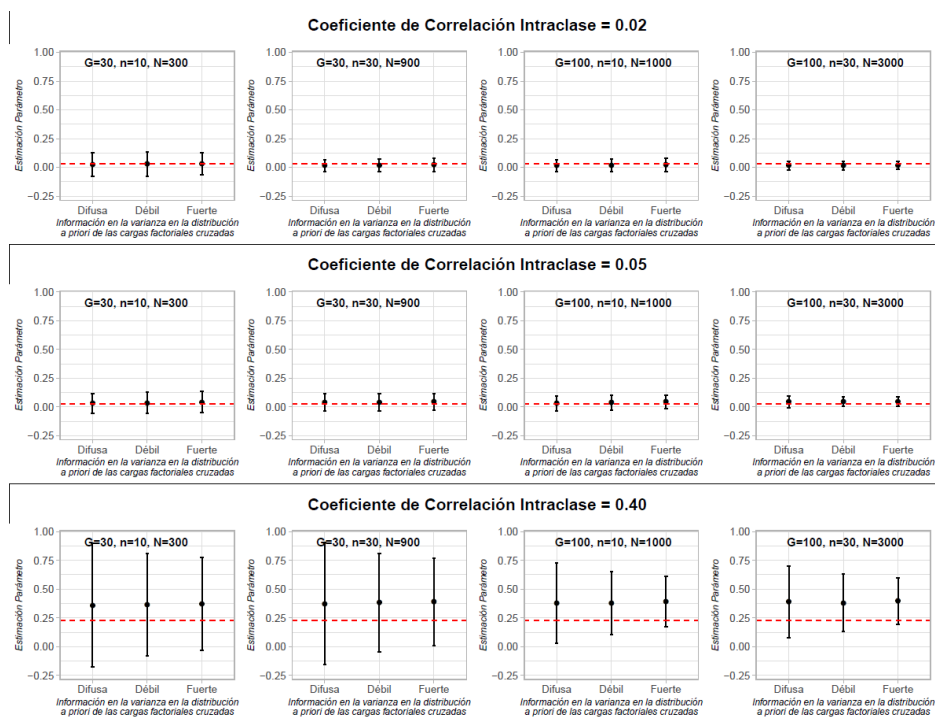


Gráfico 53 **Varianzas residuales, escenario 2:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

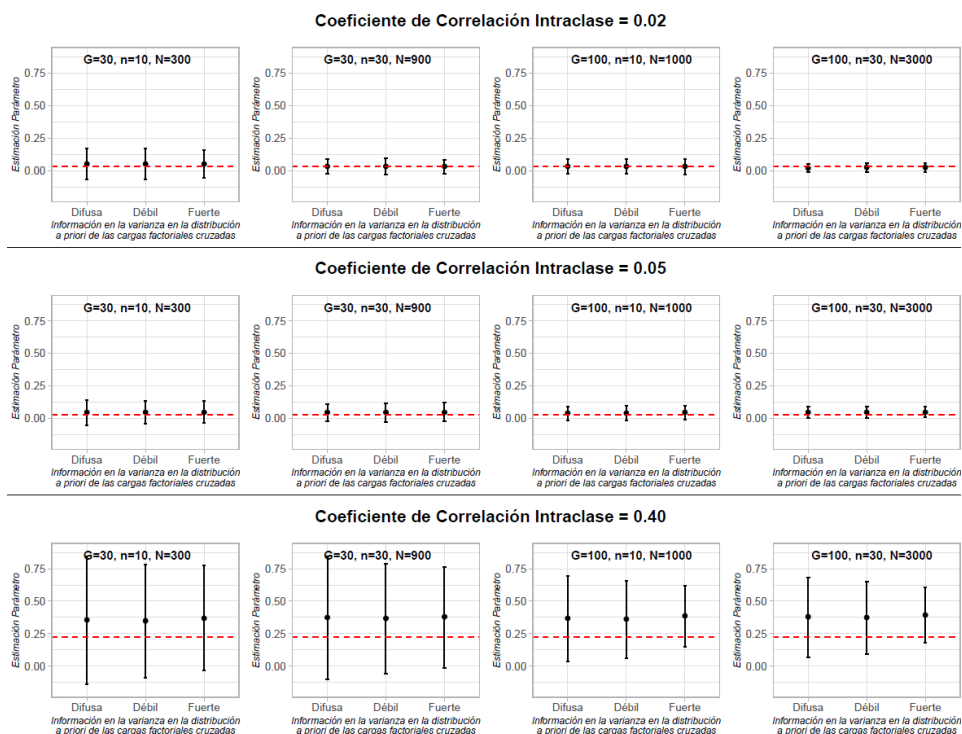


Gráfico 54 **Varianzas residuales, escenario 3:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

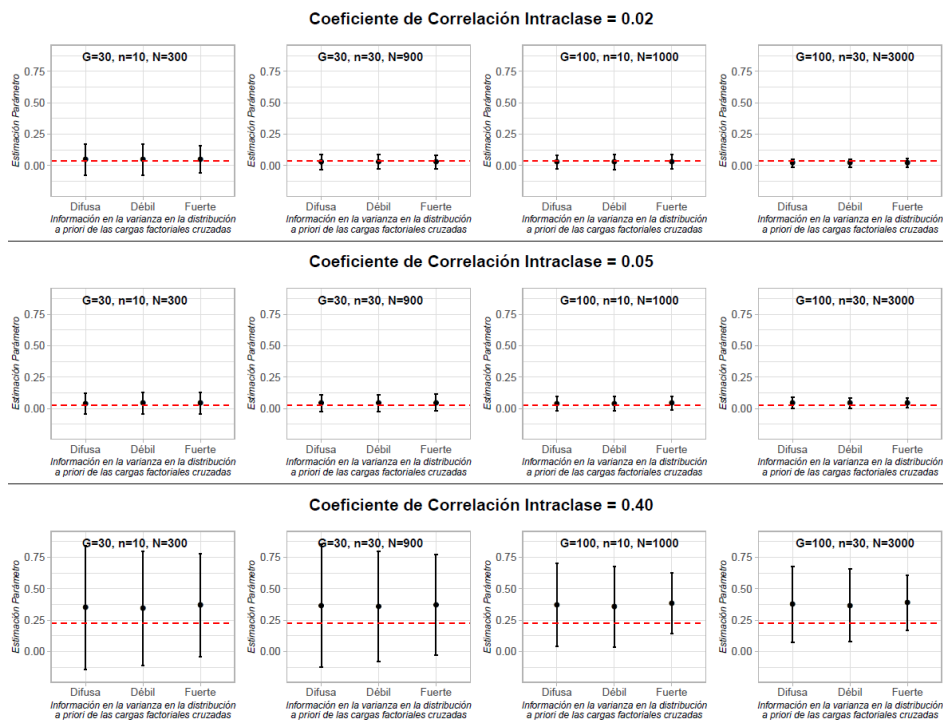


Gráfico 55 **Varianzas residuales, escenario 4:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

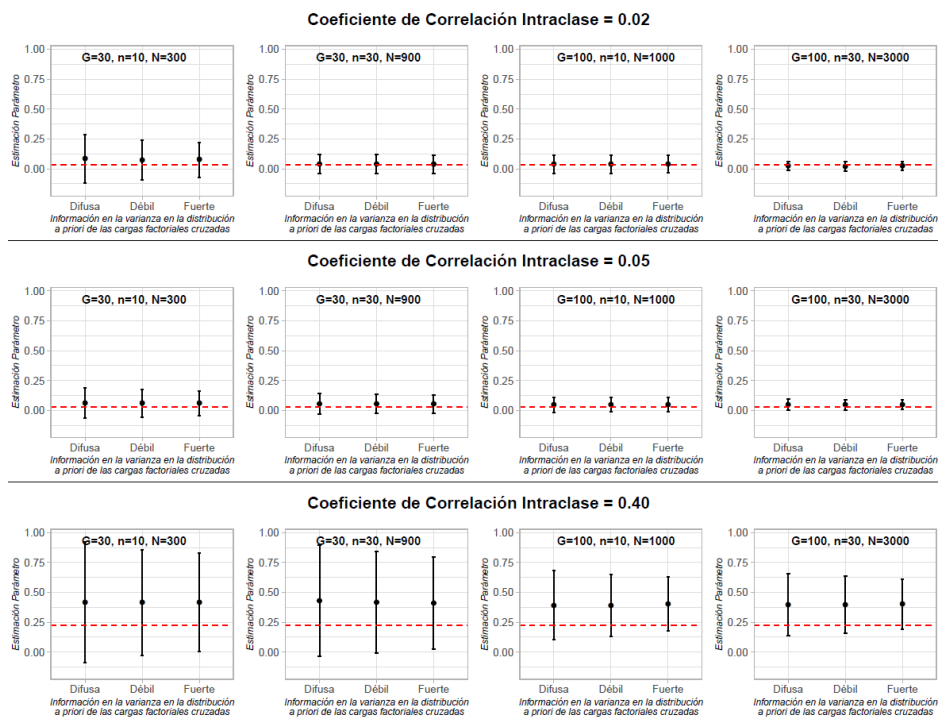


Gráfico 56 **Varianzas residuales, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

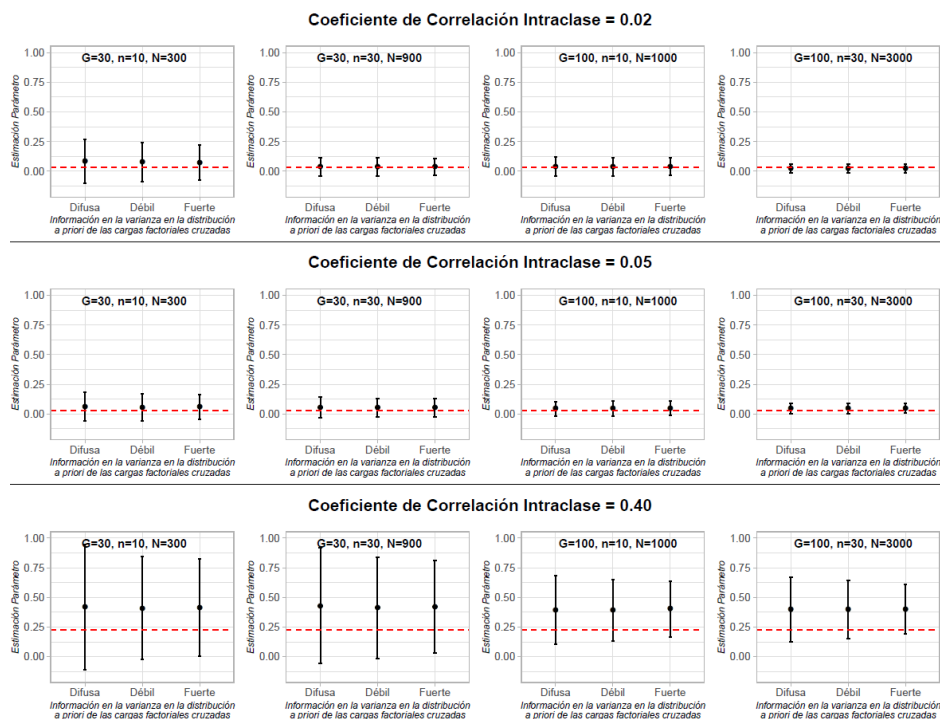


Gráfico 57 **Varianzas residuales, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclass. (Nivel entre grupos)

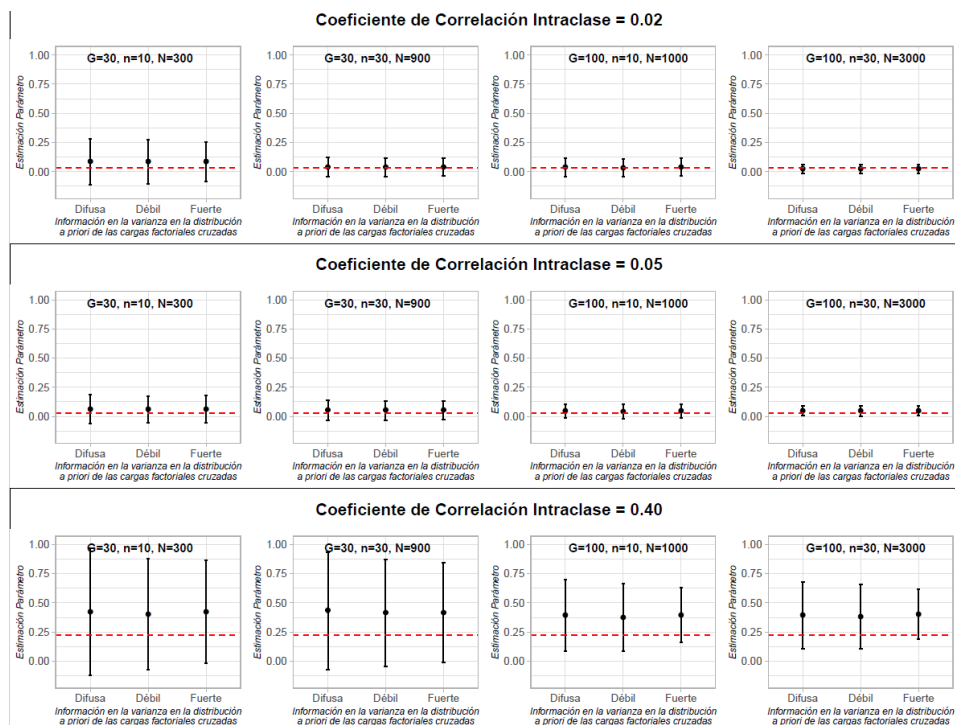


Gráfico 58 Varianza factorial, escenario 1: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

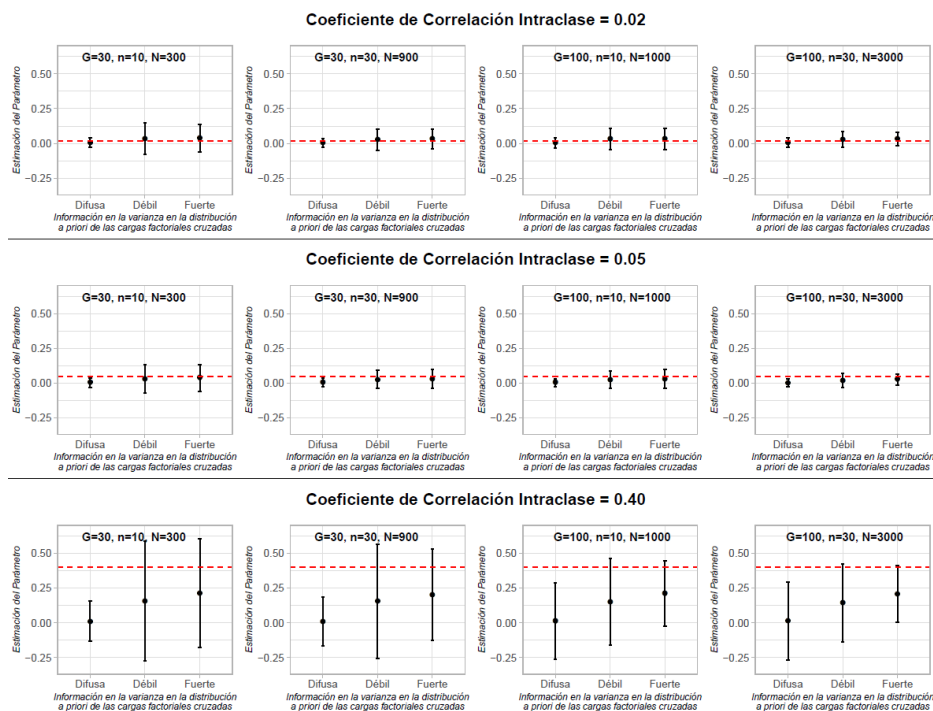


Gráfico 59 Varianza factorial, escenario 2: Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

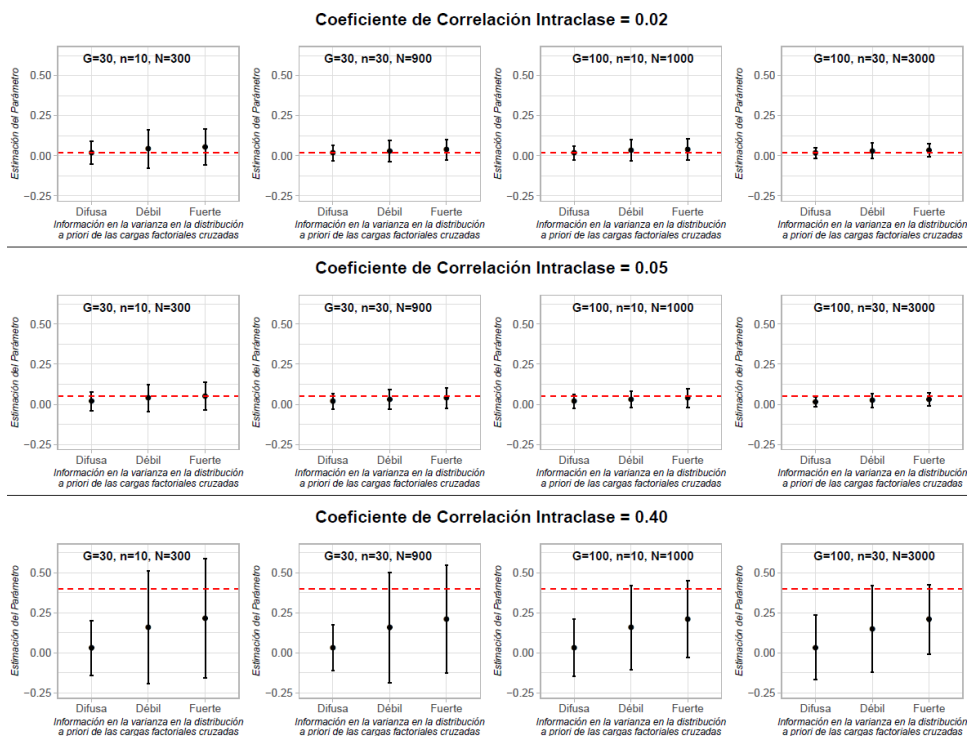


Gráfico 60 **Varianza factorial, escenario 3:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

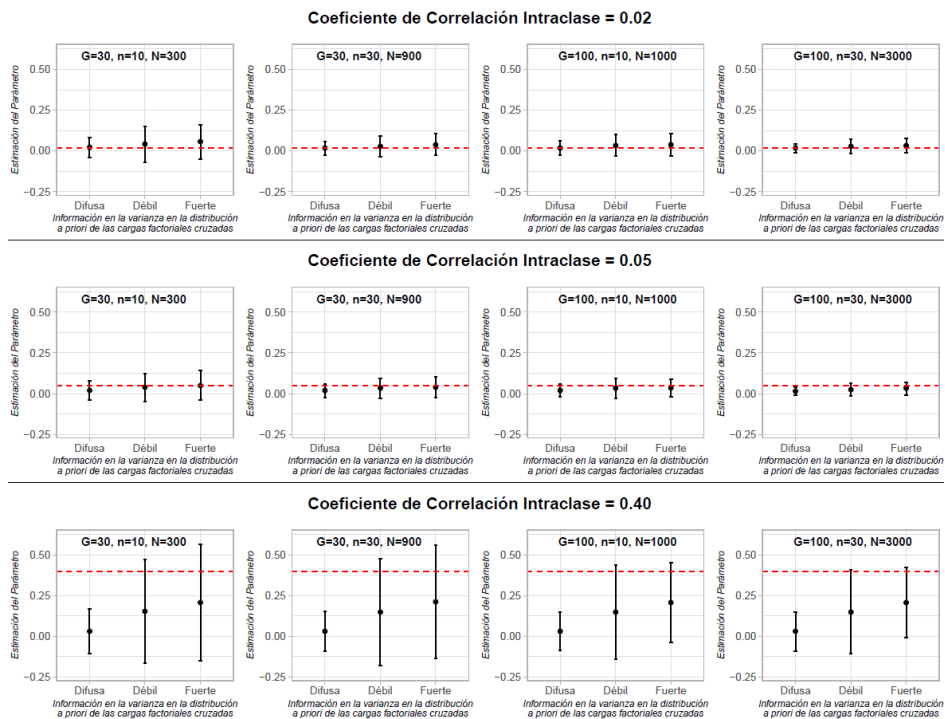


Gráfico 61 **Varianza factorial, escenario 4:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

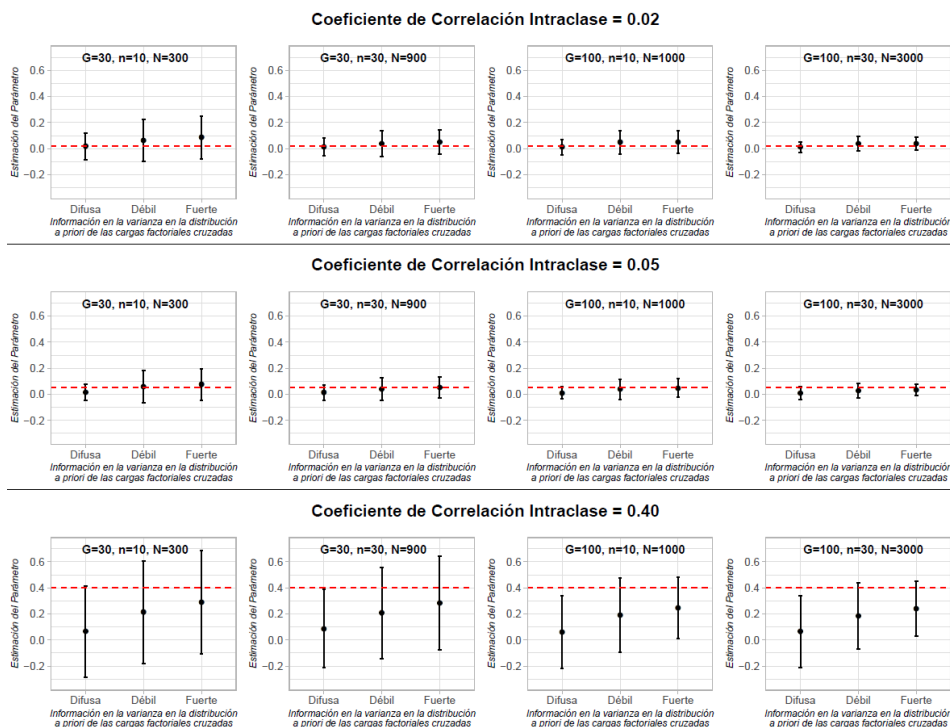


Gráfico 62 **Varianza factorial, escenario 5:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

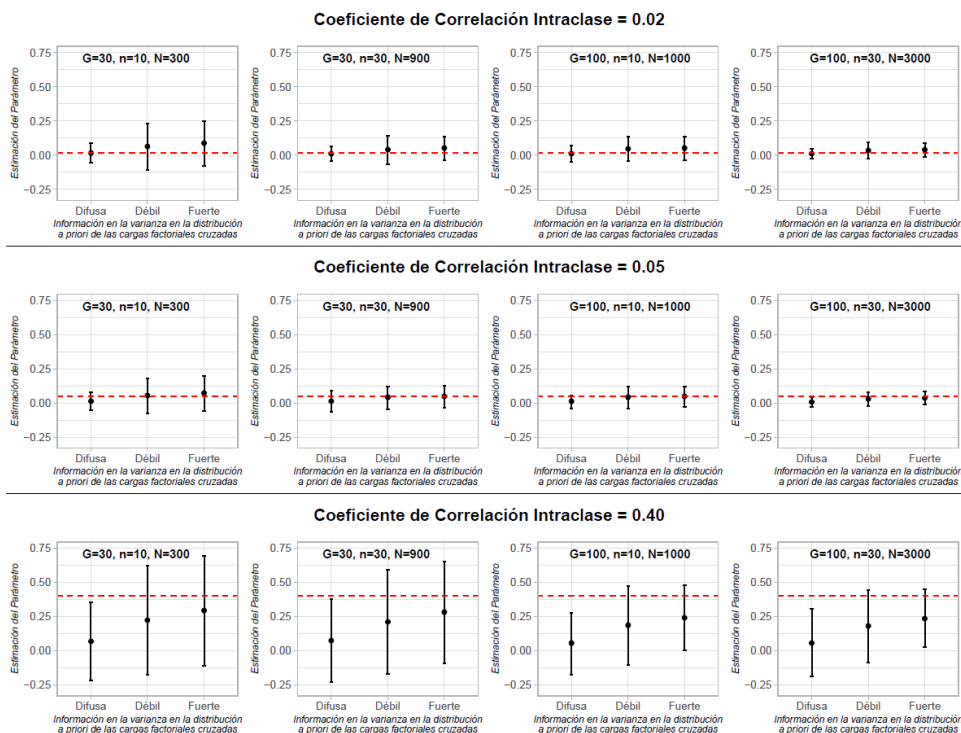


Gráfico 63 **Varianza factorial, escenario 6:** Intervalo de credibilidad de las estimaciones del parámetro, según Información en la varianza de la distribución a priori de las cargas factoriales cruzadas, Tamaño de Muestra y Coeficiente de Correlación Intraclase. (Nivel entre grupos)

