

DEPÓSITO LEGAL ZU2020000153

ISSN 0041-8811

E-ISSN 2665-0428

# Revista de la Universidad del Zulia

Fundada en 1947  
por el Dr. Jesús Enrique Lossada



**Ciencias**  
**Exactas,**  
**Naturales**  
**y de la Salud**

**78**  
**ANIVERSARIO**

**Año 16 N° 46**  
**Mayo - Agosto 2025**  
**Tercera Época**  
**Maracaibo-Venezuela**

## Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas

Jesús Eduardo Pulido \*

### RESUMEN

El objetivo de este estudio consistió en analizar la incidencia que tienen diversos tamaños de muestra en los indicadores psicométricos así como en la invarianza configural y métrica, en la escala utilizada para medir el liderazgo de la mujer en la administración educativa. Está enmarcado en el enfoque cuantitativo de tipo explicativo. Los aportes de Arias (2008), Hernández (2014), Batista y Jordi (2004) y Gaskin (2016) permitieron conformar la fundamentación teórica que orientó la investigación. El estudio se hizo con 271 mujeres y 155 hombres de República Dominicana. Los indicadores psicométricos evidenciaron validez y confiabilidad en la información dada por mujeres y hombres; los índices de bondad de ajuste concuerdan con los sugerido por Gaskin para un buen ajuste:  $\Delta CFI < 0.01$ ; por tanto, se concluye que las mujeres y los hombres entendieron de igual manera el plantemiento de los ítems utilizados en cada una de las muestras integradas.

**PALABRAS CLAVE:** Tamaño de muestra, Propiedades psicométricas, Índice de bondad de ajuste, Invarianza configural, Invarianza métrica.

\*Universidad Pedagógica Experimental Libertador (UPEL-Venezuela). ORCID: <https://orcid.org/0000-0001-6319-4533>. E-mail: [jesuspulido@upel.edu.ve](mailto:jesuspulido@upel.edu.ve)

Recibido: 15/01/2025

Aceptado: 07/03/2025

## Sample Sizes and their Impact on Invariance and Psychometric Properties

### ABSTRACT

The objective of this study was to analyze the impact that various sample sizes have on psychometric indicators as well as on configural and metric invariance, in the scale used to measure women's leadership in educational administration. It is framed in the explanatory quantitative approach. The contributions of Arias (2008), Hernández (2014), Batista and Jordi (2004) and Gaskin (2016) allowed us to form the theoretical foundation that guided the research. The study was done with 271 women and 155 men from the Dominican Republic. The psychometric indicators showed validity and reliability in the information given by women and men; The goodness of fit indices agree with those suggested by Gaskin for a good fit:  $\Delta CFI < 0.01$ ; Therefore, it is concluded that women and men understood the approach of the items used in each of the integrated samples in the same way.

**KEYWORDS:** Sample size, Psychometric properties, Goodness of fit index, Configural invariance, Metric invariance.

### Introducción

Al realizar una investigación de naturaleza cuantitativa, es necesario que se establezcan claramente las características de las unidades de población a estudiar así como la selección de una muestra representativa de la población objeto de estudio, que permita realizar análisis estadísticos; por consiguiente, el tamaño de muestra es fundamental cuando se requiere recopilar datos cuantitativos mediante instrumentos documentales para analizar sus cualidades psicométricas y la invarianza. Hernández et al. (2014) clasifica la técnica de muestreo en probabilístico y no probabilístico. En cuanto al probabilístico, Parra y Vásquez (2017) mencionan el muestreo aleatorio simple, aleatorio sistemático, aleatorio estratificado, aleatorio por conglomerados y muestro aleatorio mixto/por etapas múltiples.

Hay procedimientos apropiados para obtener tamaño de muestra, pero también se observan variados criterios para diferenciarla; es grande cuando la muestra es mayor de 30 elementos y pequeña cuando es inferior a 30 (Sarramona, 1980); si la muestra es mayor e igual a 30 se utiliza la distribución normal y si es menor e igual a 30 se utilizan las distribuciones de t

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57 de Student y Chi Cuadrado (Hamdan González, 1985; Sánchez Carrión, 1999). Cuando el tamaño de muestra es de 30 o más da una aproximación bastante buena (Weinberg y Goldberg, 1982). Se utiliza un tamaño de muestra superior a 60 cuando se hace uso de las unidades de sigma en las fórmulas de los intervalos de confianza y proporciones; si es menor de 60 las unidades de probabilidad o el número de errores estándar se deben tomar de la distribución de Student (Briones, 1998).

Pudiera pensarse que estos criterios ya han sido superados por el avance de la tecnología y la existencia de una variedad de software estadísticos, pero no es así, en el estudio realizado por Vargas y Mora (2017) sobre el tamaño de la muestra en modelos de ecuaciones estructurales con constructos latentes, hacen mención a los criterios mencionados por diversos autores, por ejemplo, (Kline, 2011; Schumacker y Lomax, 2010) sostienen que se requieren muestras grandes en los estudios de los modelos de ecuaciones estructurales (SEM, por sus siglas en inglés); pero ante esta imprecisión surge la pregunta ¿una muestra grande será la que es mayor a 30 o 60 unidades de estudio?

Vargas y Mora también hace referencia a los criterios presentados por autores como Catena et al., 2003; Hair et al., 2014; Stevens, 2009 sobre la cantidad de casos absolutos en el estudio de las SEM, sugieren 200 casos. Sin embargo, Kline (2011) advierte que una muestra de este tamaño puede resultar pequeña cuando el modelo es complejo, no existe normalidad multivariada o se utilizan ciertos tipos de estimación. Hair et al (2014) agregan que muestras muy grandes, superiores a 200, tienen el inconveniente de generar modelos muy sensibles.

En cuanto a los casos por parámetros en el estudio de las SEM, Hair et al (ob cit) recomienda un mínimo de 100 casos para modelos con 5 o menos constructos, cada uno de los factores con más de 3 indicadores y con comunalidades por ítems superiores a 0,60. Kline (2011) sugiere como muestra ideal 20 casos por parámetro y como muestra menos ideal, 10 casos por parámetro.

Con relación a los casos por variable observada Catena et al (2003, citado por Vargas y Mora) considera que 8 casos es una muestra adecuada para el total de variables observadas (ítems) y latentes. Fabrigar et al., 1999; Ferrando y Anguiano-Carrasco, 2010 citado por Lloret

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57 et al (2014), sugieren que se mantengan 3 o 4 ítems por factor, cuando el tamaño mínimo de muestra sea 200 casos.

Ante esta variedad de criterios con relación al tamaño de la muestra, en algunos casos ambiguos (muestras grandes, pequeñas y modelos complejos, cantidad de factores e indicadores por factores), pudiera generar incertidumbre y hasta confusión al momento de tomar una muestra o muestras grupales (género, estado civil, regiones, etc.) para analizarlas con los procedimientos estadísticos que correspondan, según el tipo de escala que se utilice y los objetivos del estudio que se va a realizar. Por consiguiente, este estudio está orientado a buscarle respuesta a la siguiente pregunta: ¿cómo inciden diversos tamaños de muestra, según género, en los indicadores psicométricos, de bondad de ajuste y en la invarianza configural y métrica en la escala utilizada para medir el liderazgo de la mujer en la administración educativa?

Para orientar la repuesta a esta pregunta se redactó el siguiente **objetivo**: con base en el análisis factorial conformatorio (AFC), analizar la incidencia que tienen diversos tamaños de muestra, según género, en los indicadores psicométricos, de bondad de ajuste y en la invarianza configural y métrica en la escala utilizada para medir el liderazgo de la mujer en la administración educativa.

Previo al uso de la técnica de modelaje mediante ecuaciones estructurales los investigadores, a través del AFC, deben evaluar los supuestos psicométricos de las variables observables así como los índices de bondad de ajuste y la invarianza configural y métrica. En este sentido, Bello Parias (2023), apoyándose en la data<sup>1</sup> relacionada con el estudio del uso de la herramienta excel, la cual está conformada por 43 ítems (con una escala de 1 a 5), 92 mujeres y 291 hombres y 7 factores; utilizó el AFC y el plugins HTMT + Validity CI Test (Gaskin, 2016) para averiguar si los ítems que corresponden a cada variable latente la están midiendo bien; los resultados obtenidos se observan en las tablas 1 y 2.

---

<sup>1</sup> Base de datos (92 mujeres y 291 hombres) recopilada por Gaskin, J.

Tabla 1. Indicadores de fiabilidad y validez convergente del modelo de medida.

Factores	CR	AVE
Personal	0.916	0,610
Experto	0.819	0,530
Productividad	0.936	0,746
Uso_trabajo	0.943	0,705
Gusto	0.941	0,694
Información	0.854	0,540
Apoyo	0.908	0,557

Tabla 2. Validez discriminante (HTMT) de la información dada por los encuestados en la consulta sobre el uso de excel.

	Personal	Experto	Productividad	Uso trabajo	Gusto	Información
Personal						
Experto	0,425					
Productividad	0,477	0,444				
Uso_trabajo	0,368	0,318	0,245			
Gusto	0,556	0,314	0,243	0,512		
información	0,440	0,312	0,243	0,596	0,543	
Apoyo	0,387	0,250	0,254	0,595	0,484	0,833

Los indicadores que se observan en la tabla 1 revelan que la información dada por los 383 informantes tiene validez convergente (AVE) y fiabilidad compuesta (CR o FC), según los criterios de referencia de Hair, Black, Babin y Anderson (2010, citados en Gaskin 2016):  $FC > 0,7$  y  $AVE > 0,5$ . La información de la tabla 2 indica la presencia de validez discriminante ya que ninguno de sus indicadores supera los umbrales de Henseler, Ringle and Sarstedt (2015): 0,85 para validez discriminante estricta y 0,90 para validez discriminante liberal.

Bello Parías continuó con el AFC para averiguar si el modelo en su conjunto tiene un buen ajuste; los indicadores encontrados se muestran en la tabla siguiente.

**Tabla 3.** Indicadores de bondad de ajuste

CMIN/DF	2,15
RMR	
Estandarizado	0,045
CFI	0,93
RMSEA	0,05

Los indicadores de bondad de ajuste se encuentran dentro del umbral de buen ajuste aceptable sugerido por Barrett (2007, en Medrano y Muñoz, 2017) identificados en negrilla; CMIN/DF = 2,15 < 3. Standardized RMR = 0,045 < **0,08**. CFI = 0,93 > **0,90**. RMSEAS = 0,05 < **0,08**.

Conservando la misma base de datos del estudio relacionado con el uso del Excel, Bello Parias creó en AMOS los grupos (hombres y mujeres) y al ejecutar el plugin view test (Gaskin, 2016) obtuvo los índices de bondad de ajuste que se observan en la tabla 4, los cuales indican que se está cumpliendo con la invarianza configural.

**Tabla 4.** Índices de bondad de ajuste y los criterios sugeridos por diversos autores.

	Indicador de ajuste	Criterios
CMIN/DF	1,88	< 3
RMR		
Estandarizado	0,05	< 0,08
CFI	0,91	> = 0,90
RMSEA	0,04	< 0,08

Este académico prosiguió con el estudio de la invarianza y en esta ocasión eligió en AMOS la opción multigrupo y encontró los resultados que refleja la tabla 5, en los que se observa que no hay diferencia en el CFI en los modelos, por consiguiente, hay invarianza métrica y de estructura; es decir, el modelo de medida tiene invariancia configural, métrica y de estructura.

**Tabla 5.** Indicadores RFI, TLI y CFI.

Model	RFI -rho1	TLI -rho2	CFI
Unconstrained	0,79	0,89	<b>0.90</b>
Measurement weights	0,79	0,89	<b>0.90</b>
Structural covariances	0,79	0,89	<b>0.90</b>
Measurement residuals	0,79	0,89	<b>0.90</b>

Caycho et al (2020) realizaron un estudio orientado a analizar las evidencias iniciales de validez basada en la estructura interna, la confiabilidad y la invarianza factorial entre hombres y mujeres del who-5 wbi (evaluación del bienestar) en la población universitaria de un país latinoamericano (Perú). Utilizaron un diseño instrumental, debido a que se evaluaron las propiedades psicométricas de un instrumento de medida (who-5 wbi, versión española), conformado por 5 ítems tipo Likert con 4 alternativas de respuesta (nunca, a veces, muchas veces y siempre); en el estudio participaron 228 hombres y 271 mujeres, estudiantes de universidades públicas y privadas.

La invarianza factorial entre hombres y mujeres observada es  $\Delta CFI < 0.01$ . El análisis estadístico de las respuestas dadas por los estudiantes fue hecho mediante el análisis factorial confirmatorio, porque teóricamente la escala asume la unidimensionalidad. Evaluaron el modelo original de un solo factor y los índices de ajuste no fueron del todo aceptables:  $W$  (coeficiente omega) = 0,787;  $\chi^2/gl = 7.204$ ; CFI= 0.966; RMSEA = 0.112 y SRMR = 0.032. Ante la presencia de estos índices incluyeron la corrección de errores identificados, lo que mejoró el ajuste del modelo:  $W_{\text{corregido}} = 0,758$ ;  $\chi^2/gl = 2.416$ ; CFI = 0.994; RMSEA = 0.053 y SRMR = 0.018. Además, reportaron que la cargas factoriales para este modelo oscilaron entre 0,814 para ítem 1 y 0,625 para el ítem 5.

Por tanto, estos resultados apoyan la estructura unidimensional del who-5 wbi, adecuada confiabilidad e invarianza factorial en función del sexo, lo cual indica que es un instrumento adecuado para la evaluación del bienestar.

## 1. Perspectiva teórica

Para orientar y fundamentar teóricamente este estudio se revisó la finalidad del análisis factorial confirmatorio, la normalidad multivariada así como las propiedades psicométricas de las escalas, los índices de bondad de ajuste y la invarianza factorial.

### 1.1. Análisis factorial confirmatorio (AFC)

Según Arias (2008) el análisis factorial se refiere a un grupo de técnicas de análisis de datos diseñadas para identificar factores o dimensiones inobservables empíricamente y que pueden explicar las relaciones que existen entre un grupo de variables observadas. Prosigue el autor, cada uno de esos factores o dimensiones constituyen un constructo y las variables observadas que están relacionadas con él son sus indicadores.

El propósito del análisis factorial confirmatorio (AFC) se centra en el estudio de los modelos de medida, es decir, permite analizar las relaciones entre un conjunto de indicadores o variables observadas y una o más variables latentes o factores. Los indicadores pueden ser, por ejemplo, los ítems de un test, las puntuaciones obtenidas por los sujetos en distintas escalas o los resultados provenientes de instrumentos de clasificación conductual.

### 1.2. Normalidad multivariada

La distribución normal multivariada es uno de los supuestos fundamentales en el análisis estadístico paramétrico, tales como el análisis de varianza (ANOVA), el método de máxima probabilidad o de máxima verosimilitud (ML), correlación canónica, regresión lineal múltiples, entre otros. Es importante verificar el cumplimiento de este supuesto para que los resultados obtenidos con los métodos mencionados tengan validez (Porras, 2016). La evaluación de la normalidad multivariada se puede hacer con programas como R o AMOS y con excel, se realiza la prueba Mardia.

### 1.3. Propiedades psicométricas de las escalas

Las escalas, como la de tipo Likert, se utilizan para cuantificar atributos, cualidades, constructos o conceptos teóricos no observables directamentes, que no pueden ser medidos de otra manera. La información recolectada con los instrumentos de medición debe ser válida y confiable, es decir, mostrar altos valores de validez y confiabilidad (Hernández Sampieri et al, ob cit); según estos autores:

la validez se refiere al grado en que un instrumento mide realmente la variable que pretende medir. Por ejemplo, un instrumento válido para medir la inteligencia debe medir la inteligencia y no la memoria. La confiabilidad se refiere al grado en que su aplicación repetida al mismo individuo u objeto produce resultados iguales. (p. 200).

En la práctica cuando se utilizan escalas para medir constructos (variables no observables) es necesario conocer la validez convergente (AVE) y la discriminante; la primera se da cuando se miden constructos similares y las puntuaciones que ofrecen están relacionadas empíricamente. Un test presenta validez discriminante cuando las puntuaciones que se obtienen no están relacionadas con las que ofrecen tests que supuestamente miden constructos distintos y, por tanto, se refiere al grado en que ambas medidas son empíricamente independientes (Talledo y Sánchez, 2007); dicho con palabras de Batista-Foguet y Jordi Alonso (2004) “La validez convergente se refiere a que las medidas de un mismo concepto deben estar relacionadas, y deben estarlo más que las medidas de conceptos distintos, lo que constituye la validez discriminante”.

Al realizar un AFC es absolutamente necesario establecer la validez convergente y discriminante así como la confiabilidad (Gaskin, ob cit). Según este autor existen algunas medidas que son útiles para valorar la validez y la confiabilidad: confiabilidad compuesta ( $CR > 0,7$ ), varianza promedio extraída o validez convergente ( $AVE > 0,5$ ). Según Salessi y Omar (2018) los valores de varianza media extraída, de su raíz cuadrada y de la proporción heterorasgo-monorasgo (HTMT) indican capacidad de *discriminación* entre los constructos ( $AVE > 0,50$ ;  $\sqrt{AVE} < r$ ;  $HTMT < 0,90$ ).

#### 1.4. Índices de bondad de ajuste

Como se ha dicho anteriormente el AFC permite evaluar la validez y la fiabilidad de cada ítem del cuestionario y a su vez permite realizar un contraste de hipótesis con la relación entre indicadores y dimensiones latentes. De esta forma, analizando la estructura interna del mismo, se puede verificar que el instrumento realmente evalúe el constructo que dice medir y se prueba si los datos empíricos recogidos por el instrumento se ajustan al modelo teórico que subyace al mismo; este ajuste se evalúa mediante los índices absoluto e incrementales (Muiños y Federico, 2021).

Gaskin, (2016), plantea que “la bondad del ajuste está inversamente relacionada con el tamaño de la muestra y el número de variables en el modelo”. Sugiere como guía los umbrales que están en la siguiente tabla

Tabla 6. Indicadores de bondad de ajuste.

Medida	Límite
chi-cuadrado/gl (cmin/gl)	< 3, Bueno; < 5, A veces permitido.
Valor p para el modelo	> 0,05
CFI	> 0,95, Excelente; > 0,90, Tradicional; > 0,80. A veces permitido.
GFI	> 0,95
AGFI	> 0,80
SRMR	< 0,09
RMSEA	< 0,05, Bueno; 0,05 - 0,10, Moderado; > 0,10, Malo.
PCLOSE	> 0,05

#### 1.5. Invarianza

Para llevar a cabo una comparación adecuada del concepto de interés entre diferentes grupos, es necesario que este concepto de interés se entienda de la misma manera en los diferentes grupos. Para asegurar una comparabilidad adecuada, es necesaria la invarianza de medida (Ariely y Davidov, 2012; Davidov et al., 2011; Rees-kens y Hooghe, 2008, en Coromina, 2015).

Según Caromina, la invarianza de medida es importante, dado que si no se tiene en cuenta se puede obtener una representación inexacta de las comparaciones de los conceptos de interés entre los grupos. Por ejemplo, las medias tradicionales basadas en las puntuaciones compuestas miden el concepto de interés con un único ítem o agregación de ítems y no tienen en cuenta la invarianza de medida.

## 2. Abordaje metodológico

Analizar la incidencia que tienen diversos tamaños de muestra no probabilística, según género, en los indicadores psicométricos, de bondad de ajuste y la invarianza configural y métrica, está enmarcado en el enfoque cuantitativo de tipo explicativo, porque la finalidad es evaluar la relación de causa y efecto entre los tamaños de muestra y los indicadores psicométricos, de bondad de ajuste e invarianza. “La investigación explicativa intenta ir más allá de lo exploratorio y descriptivo de investigación para determinar las causas reales que produce un fenómeno” (Mousalli, 2015, p.21).

Este estudio se realizó con el apoyo del AFC y de los resultados obtenidos en la escala utilizada para medir el efecto del liderazgo de la mujer en la administración de los centros educativos de República Dominicana<sup>2</sup>. Esta variable fue operacionalizada y surgieron las dimensiones (variables latente) que se presentan en la siguiente tabla 7 con sus respectivos ítems.

Con los 76 ítems se elaboró una escala tipo Likert con 5 alternativas de respuesta (nada, poco, suficiente, bastante y mucho), la cual fue respondida, en mayo de 2022, por 271 mujeres y 155 hombres que se desempeñaban como docentes, en 3 distritos escolares, en los niveles educativos: inicial, primaria y secundaria, en República Dominicana.

---

<sup>2</sup> Base de datos (271 mujeres y 155 hombres) recopilada por el Prof. Paula Soto, G.

Tabla 7. Distribución de ítems por dimensiones

Dimensiones/Variables latentes	Ítems
Apoyo a la calidad docente (ACD)	1 a 17
Gestión estratégica de recursos (GER)	18 a 24
Colaboración más allá de la escuela (CE)	25 a 31
Fijación y evaluación de metas educativas (FEM)	32 a 40
Capacidad para compartir el liderazgo (CCL)	41 a 45
Formación en habilidades pedagógicas (FHP)	46 a 54
Convertir la dirección en una profesión atractiva (DPA)	55 a 64
Apoyo a las labores pedagógicas de la dirección (ALPD)	65 a 76

## 2.1. Procedimientos

Para analizar la incidencia que tienen diversos tamaños de muestra, según género, en los indicadores psicométricos así como los de bondad de ajuste y la invarianza, en la escala utilizada para medir el liderazgo de la mujer en la administración educativa, se utilizó varias fases, en las cuales se incluyó diferentes tamaños de muestra para mujeres y hombres, elegidas aleatoriamente. En la primera fase (AFC1), de 271 mujeres se eligió de manera intencional, con el apoyo de SPSS 27, 95 mujeres (35,06%) y de 155 hombres se eligió 120 (77,42%); para cada una de estas muestras, con sus 8 dimensiones (variables latentes/constructos/factores), 76 ítems, el programa AMOS (versión 24) y el plugin CFA Auto Modeler (Gaskin y Lim., 2016) se construyó el correspondiente modelo de medida (gráfico de senderos) y se verificó la presencia de normalidad multivariada, con el programa AMOS; este resultado se contrastó con el criterio de Bollen (1989, citado en Nuviala, et al, 2014):  $p(p + 2)$ , siendo  $p$  el número de variables observadas.

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57

Con base en el plugin HTMT + Validity CI Test (Gaskin, ob cit) se obtuvo los indicadores psicométricos (validez convergente, discriminante y fiabilidad compuesta). Estos indicadores se evaluaron a la luz de los criterios de Gaskin para determinar si la información dada por las **95 mujeres** y los **120 hombres** confirmaba la validez y confiabilidad; la fiabilidad compuesta (CR) debe ser de 0,70 o más. Cuando la validez convergente (AVE) es menor de 0,5, la salida del plugin indica la o las variables observadas que se deben sacar del análisis y al correr nuevamente el plugin la validez convergente de cada factor así como el índice de fiabilidad compuesta (IFC) aumenta considerablemente (AVE  $\geq$  0,5; IFC o CR  $\geq$  0,7). En cuanto a la validez discriminante la salida de este plugin está fundamentada en un nuevo criterio, el cual hace referencia a la Relación Heterotrait-Monotrait (HTMT) entre los factores; los umbrales para su interpretación son: 0,85 para la validez estricta y 0,90 para la validez discriminante liberal. (Henseler, Ringle y Sarstedt 2015)

Una vez excluidas del análisis las variables observables en la muestra de mujeres y hombres se integró una sola data y mediante el análisis factorial confirmatorio (AFC), se observó la calidad de los índice de validez convergente, discriminante y fiabilidad de los ítems de cada factor (variable latente). Además, se observó la presencia de normalidad multivariada lo que permitió utilizar el método de máxima verosimilitud (ML, por sus siglas en inglés); se analizó el ajuste global del modelo de medida (gráfico path) y, tomando en cuenta el género, se examinó la invarianza configural y métrica (Multiple - Group Analysis- AMOS 24).

Hubo la necesidad de hacer varias reespecificaciones al modelo de medida para obtener los índices de bondad de ajuste (cmin/gl, SRMR, CFI y RMSEA ), estos valores se ilustran en el apartado denominado **análisis e interpretación de resultados**.

Todo este proceso se repitió en cada una de las 3 fases restantes: 90 hombres y 110 mujeres (AFC- F2), 118 mujeres y 98 hombres (AFC- F3) y 140 mujeres y 155 hombres (AFC - F4).

### 3. Análisis e interpretación de resultados

Los resultados tienen que ver con la normalidad multivariada, el análisis factorial confirmatorio, propiedades psicométricas, índices de bondad de ajuste y la invarianza configural y métrica.

### 3.1. Normalidad multivariada

En cada una de las diferentes muestras (mujeres y hombre) se conservó los 76 ítems (variables observadas) de la escala utilizada para medir el liderazgo de la mujer en la administración educativa, en tres distritos escolares de República Dominicana; por lo tanto, los resultados de asimetría y kurtosis originados por el test de normalidad multivariante en AMOS son menores a 5928, valor obtenido con la fórmula de Bollen (ob cit),  $[76(76 + 2)]$ , lo que evidencia normalidad multivariada, tal como se observa en la tabla 8.

Tabla 8. Normalidad multivariada, según género.

Normalidad multivariante AMOS				
	Tamaño de muestra	Asimetría	Kurtosis cr	Bollen $p(p + 2)$
<b>Fase 1</b>				
Mujeres	95	180,685	8,087	<b>5928</b>
Hombres	120	156,231	7,859	
<b>Fase 2</b>				
Mujeres	110	234,36	11,287	<b>5928</b>
Hombres	90	68,258	2,974	
<b>Fase 3</b>				
Mujeres	118	276,401	13,787	<b>5928</b>
Hombres	98	96,493	4,386	
<b>Fase 4</b>				
Mujeres	140	357,48	19,423	<b>5928</b>
Hombres	155	242,407	13,876	

### 3.2. Propiedades psicométricas

Al ejecutar la fase 1 (AFC) para las mujeres, la primera salida del plugin HTMT + Validity CI Test generó índices de validez convergente (AVE) menores a 0,5 en los factores CE, CCL y

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57  
 DPA; en el caso de los hombres los factores con AVE menor a 0,5 fueron ACD, GER, CCL y DPA; esta primera salida indicó los ítems que debían ser excluidos del modelo de medida y al ejecutar el plugin nuevamente se obtuvo los valores reflejados en la tabla 9, los cuales evidencian validez convergente, discriminante (HTMT) y fiabilidad compuesta (FC o CR), según los criterios de Gaskin, Henseler, Ringle y Sarstedt (ob cit).

**Tabla 9.** Propiedades psicométricas según tamaños de muestra y género. Fase 1

95 Mujeres			Relación Heterotrait - Monotrait (HTMT)									Ítems excluidos
Factores	FC o CR	AVE	Factores	ACD	GER	CE	FEM	CCL	FHP	DPA	ALPD	
ACD	0.948	0.517	ACD									
GER	0.896	0.555	GER	0.797								
CE	0.813	0.524	CE	0.443	0.480							CE26 CE28 CE29
FEM	0.93	0.598	FEM	0.788	0.71	0.377						
CCL	<b>0.798</b>	0.506	CCL	0.774	0.71	0.319	0.752					CCL44
FHP	0.905	0.519	FHP	0.693	0.776	0.553	0.715	0.829				
DPA	0.887	0.522	DPA	0.191	0.371	0.615	0.217	0.28	0.44			DPA64
ALPD	0.948	0.613	ALPD	0.304	0.361	0.118	0.419	0.553	0.39	0.671		

  

120 Hombres			Relación Heterotrait - Monotrait (HTMT)									Ítems excluidos
Factores	FC o CR	AVE	Factores	ACD	GER	CE	FEM	CCL	GHI	DPA	ALPD	
ACD	0.942	0.504	ACD									ACD17
GER	0.86	0.509	GER	0.781								GER22
CE	0.883	0.522	CE	0.599	0.662							
FEM	0.917	0.551	FEM	0.705	0.732	0.602						
CCL	0.825	0.546	CCL	0.725	0.761	0.591	0.767					CCL44
GHI	0.919	0.558	GHI	0.713	0.677	0.589	0.756	0.709				
DPA	0.902	0.535	DPA	0.251	0.259	0.496	0.387	0.291	0.45			DPA64
ALPD	0.928	0.525	ALPD	0.343	0.325	0.431	0.381	0.403	0.48	0.728		

Tabla 10. Tamaños de muestra y propiedades psicométricas según género. Fase 2

110 Mujeres			Relación Heterotrait - Monotrait (HTMT)									Ítems excluidos
Factores	FC o CR	AVE	Factores	ACD	GER	CE	FEM	CCL	FHP	DPA	ALPD	
ACD	0.941	0.502	ACD									ACD17
GER	0.881	0.519	GER	0.797								
CE	0.833	0.502	CE	0.475	0.474							CE28-CE29
FEM	0.931	0.602	FEM	0.804	0.694	0.414						
CCL	<b>0.797</b>	0.502	CCL	0.769	0.717	0.423	0.711					CCL44
FHP	0.896	0.521	FHP	0.696	0.758	0.615	0.677	0.813				FHP46
DPA	0.888	0.525	DPA	0.18	0.326	0.58	0.214	0.277	0.462			DPA64
ALPD	0.943	0.589	ALPD	0.299	0.36	0.203	0.414	0.535	0.399	0.676		

  

90 Hombres			Relación Heterotrait - Monotrait (HTMT)									Ítems excluidos
Factores	FC o CR	AVE	Factores	ACD	GER	CE	FEM	CCL	FHP	DPA	ALPD	
ACD	0.942	0.509	ACD									ACD17
GER	0.845	0.524	GER	0.783								GER20-GER22
CE	0.897	0.558	CE	0.572	0.637							
FEM	0.917	0.554	FEM	0.657	0.76	0.594						
CCL	0.837	0.564	CCL	0.709	0.769	0.602	0.788					CCL44
FHP	0.922	0.568	GHI	0.676	0.701	0.603	0.754	0.728				
DPA	0.884	0.508	DPA	0.287	0.313	0.52	0.462	0.409	0.524			
ALPD	0.938	0.561	ALPD	0.295	0.393	0.406	0.365	0.439	0.446	0.699		

Tabla II. Tamaños de muestra y propiedades psicométricas según género. Fase 3

118 Mujeres			Relación Heterotrait - Monotrait (HTMT)									Ítems excluidos
Factores	FC o CR	AVE	Factores	ACD	GER	CE	FEM	CCL	FHP	DPA	ALPD	
ACD	0.942	0.506	ACD									ACD17
GER	0.887	0.532	GER	0.803								CE28- CE29
CE	0.808	0.516	CE	0.445	0.447							CCL30- CCL40- CCL45
FEM	0.936	0.62	FEM	0.8	0.726	0.362						PDA64
CCL	0.812	0.593	CCL	0.785	0.724	0.41	0.725					
FHP	0.901	0.504	FHP	0.724	0.768	0.619	0.69	0.775				
DPA	0.886	0.518	DPA	0.179	0.32	0.629	0.204	0.249	0.448			
ALPD	0.944	0.594	ALPD	0.311	0.358	0.204	0.408	0.501	0.407	0.68		

  

98 Hombres			Relación Heterotrait - Monotrait (HTMT)									Ítems excluidos
Factores	FC o CR	AVE	Factores	ACD	GER	CE	FEM	CCL	FHP	DPA	ALPD	
ACD	0.942	0.508	ACD									ACD17
GER	0.85	0.533	GER	0.777								GER20- GER22
CE	0.887	0.534	CE	0.56	0.637							
FEM	0.915	0.545	FEM	0.668	0.755	0.587						
CCL	0.842	0.573	CCL	0.7	0.764	0.58	0.779					CCL44
FHP	0.918	0.555	GHI	0.689	0.695	0.597	0.762	0.717				
DPA	0.885	0.508	DPA	0.251	0.27	0.512	0.415	0.307	0.491			
ALPD	0.936	0.554	ALPD	0.309	0.391	0.413	0.372	0.392	0.45	0.687		

Tabla 12. Tamaños de muestra y propiedades psicométricas según género. Fase 4

140 Mujeres			Relación Heterotrait - Monotrait (HTMT)									Ítems excluidos
Factores	FC o CR	AVE	Factores	ACD	GER	CE	FEM	CCL	FHP	DPA	ALPD	
ACD	0.947	0.514	ACD									
GER	0.892	0.544	GER	0.821								
CE	0.811	0.518	CE	0.423	0.462							CE26- CE28- CE29
FEM	0.932	0.607	FEM	0.799	0.751	0.362						
CCL	0.816	0.53	CCL	0.771	0.687	0.359	0.729					CCL44
FHP	0.905	0.517	FHP	0.709	0.743	0.59	0.72	0.774				
DPA	0.892	0.532	DPA	0.179	0.322	0.631	0.231	0.253	0.477			DPA64
ALPD	0.941	0.579	ALPD	0.302	0.358	0.256	0.404	0.474	0.423	0.678		

  

155 Hombres			Relación Heterotrait - Monotrait (HTMT)									Ítems excluidos
Factores	FC o CR	AVE	Factores	ACD	GER	CE	FEM	CCL	FHP	DPA	ALPD	
ACD	0.945	0.504	ACD									
GER	0.873	0.501	GER	0.733								
CE	0.865	0.52	CE	0.574	0.589							CE28
FEM	0.921	0.566	FEM	0.722	0.73	0.551						
CCL	0.828	0.55	CCL	0.714	0.731	0.59	0.749					CCL44
FHP	0.918	0.557	GHI	0.712	0.684	0.576	0.76	0.684				
DPA	0.901	0.539	DPA	0.208	0.196	0.574	0.309	0.3	0.38			DPA64
ALPD	0.921	0.522	ALPD	0.375	0.341	0.432	0.401	0.437	0.489	0.671		ALPD65

Una vez hecha la primera ejecución del plugin HTMT + Validity CI Test, las tablas 9, 10, 11 y 12 evidencian los ítems que deben excluirse del análisis factorial confirmatorio para obtener índices superiores a 0,5 y 0,70 para la validez convergente (AVE) y fiabilidad compuesta (FC o CR), respectivamente; en cuanto a la validez discriminante la correlación entre los factores no es superior a 0,85, en ninguna de las muestras. Por consiguiente, las propiedades psicométricas reflejadas en estas tablas (9 a la 12) evidencian, que al excluir del análisis las variables observadas (ítems), independientemente del tamaño de muestra, es válida y confiable la

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57  
información dada por mujeres y hombres sobre el efecto del liderazgo de la mujer en la administración de los centros educativos de República Dominicana.

### 3.3. Muestras integradas por fase

En la integración de las muestras por fases se observa en la tabla 13 la cantidad de ítems a ser tomados en cuenta para el análisis de las cualidades psicométricas, bondad de ajuste de cada modelo y la invarianza configural y métrica.

Otro aspecto que se destaca en la tabla 13 es que la diferencia entre los tamaños de muestra de mujeres y hombres oscilan entre 15 y 25, lo que se consideró adecuado para este estudio, aunque discrepa un poco con lo planteado en el vídeo realizado por Bello Paria (ob cit), en el cual se observa una diferencia de 199 (92 mujeres - 291 hombres); este autor sostiene que la diferencia entre grupos no debe ser tan grande, pero cuando las cargas son altas no ve diferencias; no obstante, considera que no deben haber grupos pequeños, por ejemplo, 15, 40 o 50; debe haber un mínimo 100 en cada grupo y concluye: “sigue el problema de la ambigüedad en el tamaño de muestra, no hay una regla general de concenso en la comunidad científica, pero se sugiere valores no tan alejados entre los grupos”.

Tabla 13. Distribución de ítems por muestras integradas y diferencia entre los tamaños de muestra.

Fases	Tamaños de Muestra		Muestras Integradas	$\Delta$ Muestras	Cantidad de ítems	Variables Latentes - Factores
	Mujeres	Hombres				
1	95	120	215	25	71	8
2	110	90	200	20	70	8
3	118	98	216	20	67	8
4	140	155	295	15	71	8

### 3.4. Invarianza configural y métrica

Por lo general el proceso para evaluar las propiedades psicométricas de la información recolectada con un instrumento documental culmina con lo reportado en las tablas de la 9 a la 12, pero se consideró conveniente conocer si las mujeres y los hombres, con niveles de conocimientos similares sobre cada una de las 8 variables latentes, tienen la misma puntuación cruda esperada en la medición realizada sobre el liderazgo de la mujer en la administración de los centros educativos de República Dominicana.

Previo al análisis de la invarianza, con AMOS 24. se hizo el AFC en cada muestra integrada (modelo de medida) y se obtuvo las siguientes propiedades psicométricas en cada una de las fases: validez convergente (AVE) superior a 0,5, la relación **Heterotrait – Monotrait** entre los factores, menor a 0,80 (validez discriminante) y la fiabilidad compuesta, superior a 0,70; estos resultados evidencian que los ítems propuestos para medir la correspondiente variable latente tiene un grado de certeza aceptable; que no están relacionados los ítems de factores distintos, por ejemplo, no hay relación entre los ítems del factor **apoyo a la calidad docente (ACD)** con los del factor **gestión estratégica de recursos (GER)** y que hay consistencia interna en la escala.

Con base en la modificación de índices y diversas reespecificaciones realizadas a los 4 modelos (muestras integradas) se obtuvo los indicadores de bondad de ajuste que se observan en la tabla 14.

Tabla 14. Índices de bondad de ajuste, según muestras integradas

Modelos	Muestras integradas	Ítems por muestra	Índices de bondad de ajuste			
			CMIN/GI	RMR Standarizado	CFI	RMSEA
1	215	36	1.898 < 3	0.065 < 0,09	0.901	0.065 < 0,08
2	200	38	2,03 < 3	0.056 < 0,09	0,90	0.071 < 0,08
3	216	39	1.949 < 3	0.068 < 0,09	0.886 < 0,90	0.066 < 0,08
4	295	38	2.107 < 3	0.066 < 0,09	0.909	0.061 < 0,08

Según el criterio de Gaskin (ob cit) sobre los índices de bondad de ajuste (ver tabla 6), la información refelejada en esta tabla 14 revela que hay un buen ajuste global en los modelos, aunque el CFI del modelo 3 no alcanzan el umbral 0.90, pero están muy cerca.

Otro aspecto importante que se observa en esta tabla es que después de las distintas reespecificaciones hechas en cada modelo (fundamentadas sólo en el criterio estadístico) la cantidad de ítems señalados en la tabla 13 se redujo considerablemente (51%, 54%, 58% y 54%, respectivamente).

### 3.5. Invarianza configural

Una vez creado los grupos (mujeres y hombres) en cada modelo se obtuvo los indicadores de la tabla 15.

**Tabla 15.** Índice de bondad de ajuste de la invarianza configural, según muestras integradas

Modelos	Muestras integradas		IFI		TLFI	
	CMIN/DF	Delta2	rho2	CFI	RMSEA	
1	215	1,719	0,855	0,835	0,851	0,058
2	200	1,798	0,829	0,806	0,825	0,063
3	216	1,696	0,844	0,824	0,84	0,057
4	295	1,766	0,881	0,866	0,879	0,051

De la información de la tabla 15 se deduce que los índices CMIN/DF y del error cuadrático medio de aproximación (RMSEA) son adecuados porque están dentro de los linderos sugeridos; en cuanto al indicador de ajuste incremental (CFI) es menor a 0,90 en cada modelo y son similares a los IFI y a los TLFI, por consiguiente, se considera que cumple con la invarianza configural. Es decir, que el gráfico path, representado en el AFC que se construyó en cada modelo logró un ajuste adecuado en el grupo de mujeres y de hombres.

Tabla 16. Índice de bondad de ajuste de la invarianza grupal, según muestras integradas (F1 a F4)

Modelo	CMIN/GL	RMSEA	IFI Delta2	TLI rho2	CFI	F1 ΔCFI	CMIN/GL	RMSEA	IFI Delta2	TLI rho2	CFI	F2 ΔCFI
Sin restricciones	1.719	0.058	0,855	0,835	0,851		1.798	0.063	0,829	0,806	0,825	
Pesos de medición	1.727	0.058	0,849	0,833	0,846	<b>0.005</b>	1.779	0.063	0,828	0,811	0,825	<b>0</b>
Covarianzas estructurales	1.694	0.057	0,847	0,842	0,846	<b>0</b>	1.754	0.062	0,828	0,817	0,826	<b>0,001</b>
Residuos de medición	1.685	0.057	0,844	0,845	0,844	0,002	1.755	0.062	0,821	0,817	0,821	0,005
<b>Muestra Integrada = 215</b>							<b>Muestra Integrada = 200</b>					
Modelo	CMIN/GL	RMSEA	IFI Delta2	TLI rho2	CFI	F3 ΔCFI	CMIN/GL	RMSEA	IFI Delta2	TLI rho2	CFI	F4 ΔCFI
Sin restricciones	1.696	0.057	0,844	0,824	0,84		1.766	0.051	0,881	0,866	0,879	
Pesos de medición	1.692	0.057	0,841	0,825	0,838	<b>0,002</b>	1.76	0.051	0,879	0,867	0,877	<b>0,002</b>
Covarianzas estructurales	1.673	0.056	0,84	0,83	0,838	<b>0</b>	1.752	0.051	0,876	0,868	0,875	<b>0,002</b>
Residuos de medición	1.664	0.056	0,836	0,832	0,836	0,002	1.749	0.051	0,872	0,869	0,872	0,003
<b>Muestra Integrada = 216</b>							<b>Muestra Integrada = 295</b>					

### 3.6. Invarianza métrica

Una vez probada la invarianza configural en las diferentes muestras integradas, según género, se procedió a probar la invarianza métrica y para ello se utilizó la opción multigrupo (Amos 24) y se obtuvo la información que se observa en la tabla 16; un modelo sin restricciones, otro modelo donde coloca restricciones a los pesos y a las covarianzas estructurales; un tercer modelo con restricciones en los pesos de medición, en las covarianzas estructurales y en los residuos de medición.

Según los resultados de esta tabla en cada muestra integrada los indicadores de CMIN/GL y los de RMSEA están dentro de los sugeridos por Gaskin (ob cit) y las diferencias del indicador de ajuste incremental (CFI) es menor a 0,01 (Cheung y Rensvold, 2002, citado en De Francisco et al, 2020), lo que evidencia presencia de invarianza métrica y la estructural, por tanto, es obvio que las mujeres y los hombres entendieron de forma igual los planteamientos en los ítems de cada muestra integrada (36, 38, 39 y 38 ítems, respectivamente).

## Conclusiones

El análisis factorial confirmatorio y la respuesta dada a 76 ítems (escala tipo Likert) por diversas muestras aleatorias tomadas de 271 mujeres y 155 hombres permitió obtener las propiedades psicométricas en cada muestra, según género; este procedimiento, por lo general, no se hace, se integran en una sola data a las mujeres y hombres y se obtienen de manera global los indicadores de validez convergente, discriminante y de consistencia interna de la puntuación del instrumento, lo que dificulta tener una información desagregada (hombres y mujeres) en cuanto a estas cualidades de la información recolectada con el instrumento documental (escala). En el reporte de las tablas 9 a la 12 se evidencia la presencia de validez y confiabilidad de la información dada por las muestras de mujeres y hombres.

Tomando en cuenta la distribución multivariada de la escala, el método de estimación Maximum Likelihood (máxima verosimilitud, ML), el AFC y diversas reespecificaciones se obtuvo los índices de bondad de ajuste ilustrados en la tabla 14, los cuales reflejan un buen ajuste global en los cuatro modelos; no obstante, el indicador de ajuste incremental (CFI) correspondiente a la muestra de mayor tamaño (modelo 4) es ligeramente superior (0,909) y el indicador RMSEA (0,061) presenta poca diferencia con respecto a los demás. En este caso pareciera que el tamaño de la muestra tuvo alguna repercusión en estos indicadores.

El otro aspecto que llama la atención en esta tabla es la disminución de ítems, más del 47%, en cada una de las muestras integradas; esta reducción es producto de las variadas reespecificaciones en las cuales se eligió el valor más alto en la modificación de índices (MI) ya que ese valor permite disminuir el Chi- cuadrado al eliminar el ítem que corresponda y, por tanto, mejoran los indicadores CEMIN/GL, CFI, RMR estandarizado y el RMSEA; en este estudio la eliminación de los ítems se hizo tomando en cuenta solamente el criterio estadístico y no el teórico, lo conveniente es tomar en cuenta los dos criterios. Este deber ser (criterio teórico y estadístico) conlleva a pensar qué hacer cuando el ítem o los ítems teóricamente no deben ser eliminados pese a que su índice de bondad de ajuste no es bueno, es decir, no está en consonancia con los criterios existentes.

En cuanto a la invarianza configural, la información de la tabla 15 revela que en las diferentes muestras integradas el indicador de ajuste incremental (CFI) es superior a 0,80 e

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57 inferior a 0,90; esto puede dar lugar a pensar que no cumple con esta invarianza, pero Gaskin (2016) considera que a veces este indicador es permitido cuando es mayor a 0,80 y en este caso la mayoría de las cargas factoriales de los ítems de cada factor son superiores a 0,70, el mismo caso se presenta con el CFI ilustrado en la tabla 16, pero su diferencia entre los modelos es menor a 0,01 lo que evidencia la presencia de invarianza configural y métrica, por consiguiente, se deduce que las mujeres y los hombres entendieron de igual manera lo planteado en los ítems que permitieron medir el liderazgo de la mujer en la administración educativa.

En consecuencia, la comprobación de la invarianza debe ser un complemento de las propiedades psicométricas en las escalas porque además de conocer la validez y confiabilidad de la información dada por los informantes se logra saber si los diferentes grupos (género, regiones, nivel académico, edad, etc.) han entendido de igual forma los reactivos de una escala, diseñada para recolectar información sobre un determinado constructo psicológico. Podría darse el caso de recolectar información válida y confiable y no todos los informantes de los grupos haber interpretado de la misma forma los ítems de una o varias variables latente; esto no es deseable porque pudiese tener alguna repercusión en los resultados de análisis estadísticos posteriores, como la inferencia en estudios comparativos.

Otro de los aspectos señalados al inicio de este estudio es la disparidad de criterios con respecto al tamaño de la muestra: 5 y 10 casos por ítems, 10 o 20 casos por parámetro; si se elige 5 casos por ítem para los 76 ítems de la escala que se ha venido analizando se necesitaría una muestra de 380 casos ( $76 * 5$ ) y se toma el criterio de 10 casos por ítem la muestra debe tener 760 informantes. En cuanto a las muestras integradas para los 39 ítems (modelo 3, tabla 14), por ejemplo, se necesitaría un mínimo de 195 ( $5 * 39$ ) y un máximo 390 casos ( $39 * 10$ ).

Los tamaños de muestra no sólo deben depender de la cantidad de ítems del instrumento y de los criterios existentes sino también de la carga factorial entre el ítem y el factor o constructo, por ejemplo, para cargas factoriales de 0,5, 0,6, 0,7 y 0,75 se necesita un tamaño de muestra de 150, 85, 60 y 50, respectivamente (Hair y Colds, 1999, en Roco Videla et al, 2021).

Por consiguiente, los diversos tamaños de muestras utilizados para este estudio no afectaron la buena validez y confiabilidad en la información dada por mujeres y hombres; de igual modo no se observa diferencias importantes en los indicadores de bondad de ajuste en los

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57  
cuatro modelos analizados y se verificó la invarianza configural y métrica; estos resultados favorables probablemente se deban, en parte, a que la mayoría de las cargas factoriales en las variables observables (ítems) son mayores e iguales a 0,70.

Al examinar los resultados del estudio de Bello Paria y (2023), conformado por 43 ítems, 7 factores, 92 mujeres y 291 hombres evidencia validez y confiabilidad, índices de bondad de ajuste adecuados ( $CFI \geq 0,90$ ), presencia de invarianza configural y métrica. El estudio realizado por Caycho-Rodríguez et al (2020) estuvo integrado por 5 ítems tipo Likert, 1 factor, 228 hombres y 271 mujeres, también reporta validez y confiabilidad en la información dada por los estudiantes de universidades públicas y privadas, presenta buenos indicadores de bondad de ajuste ( $CFI \geq 0,96$ ) y está presente la invarianza configural y métrica.

Como se puede observar, la diferencia entre estos dos estudios y el desarrollado en 4 fases está en los tamaños de muestra y en el indicador de ajuste incremental ( $0,80 < CFI < 0,90$ ), pero como se ha indicado anteriormente la decisión de tomar el CFI como buen ajuste de los modelos con su respectiva muestra se debe a las cargas factoriales antes mencionadas ( $> = 0,70$ ) y al criterio de Gaskin, con relación al  $CFI > 0,80$ .

Si el propósito de este estudio hubiese sido utilizar una familia de modelos estadísticos multivariantes que permitan estimar el efecto y relaciones entre múltiples variables, relacionadas con el liderazgo de la mujer en el ámbito educativo, probablemente los tamaños de muestra utilizados en las 4 fases señaladas generarían resultados de credibilidad dudosa; por ello, el tamaño de la muestra debe tomar en cuenta el alcance del estudio que se piensa llevar a cabo así como los resultados que se obtenga del pilotaje de la escala o escalas a ser utilizadas para recolectar información sobre un constructo psicológico.

-Limitaciones en la búsqueda de la incidencia de diferentes tamaños de muestra en las propiedades psicométrica, índice de bondad de ajuste e invarianza factorial

Por lo general, en las investigaciones que he realizado quedan aspectos importantes que, involuntariamente, no son tomados en cuenta durante la realización del estudio; en este caso el estudio no se hizo con más de dos grupos, no se tomó en cuenta las cargas factoriales

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57 comprendidas entre 0.5 y 0,67; no se comparó los resultados obtenidos en AMOS con los que proporciona el SmartPLS, el STATA y el JASP. Probablemente, estos resultados ayuden a despejar un poco más la incertidumbre que existe con relación al tamaño de la muestra y su incidencia en la invarianza configural y métrica como complemento de las propiedades psicométricas de las escalas de medición y de los indicadores de bondad de ajuste de los modelos de medida.

## Referencias

Arias Martínez, B. (2008). Desarrollo de un ejemplo de análisis factorial confirmatorio con LISREL, AMOS y SAS. *Publicaciones en el Instituto Universitario de Integración en la Comunidad (INICO) Universidad de Salamanca*, pp 75 - 120. Recuperado de [https://dialnet.unirioja.es/buscar/documentos?query=Dismax.DOCUMENTAL\\_TODO=DESARROLLO+DE+UN+EJEMPLO+DE+AN%C3%8ILISIS+FACTORIAL+CONFIRMATORIO+CON+LISREL%2C+AMOS+Y+SAS](https://dialnet.unirioja.es/buscar/documentos?query=Dismax.DOCUMENTAL_TODO=DESARROLLO+DE+UN+EJEMPLO+DE+AN%C3%8ILISIS+FACTORIAL+CONFIRMATORIO+CON+LISREL%2C+AMOS+Y+SAS)

Batista-Foguet, J. M. y Jordi Alonso, G. C. (2004). Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Medicina Clínica*, 122(S1), pp. 21 - 27. Recuperado de <https://www.elsevier.es/es-revista-medicina-clinica-2-articulo-analisis-factorial-confirmatorio-su-utilidad-validacion-cuestionarios-13057542>

Bello Parias, L. D. (2023, octubre 20). V16 Construcción de grupos AMOS AFC [Video]. <https://youtu.be/H60vUfiZ6r8>

Briones, G. (1998). Métodos y técnicas de investigación para las ciencias sociales. 3ª Ed. Trillas S. A. México D. F.

Caycho-Rodríguez, T., Ventura-León, J., Karla Azabache-Alvarado, K., Mario Reyes-Bossio, M. y Cabrera-Orosco, I. (2020). Validez e invarianza factorial del Índice de Bienestar General (who-5 wbi) en universitarios peruanos. *Revista Ciencias de la Salud*, 18(3). Universidad del Rosario Colombia. <https://doi.org/10.12804/revistas.urosario.edu.co/revsalud/a.9797>

Coromina, L. (2015). Importancia de la invarianza de medida de la confianza a través del tiempo. El caso de España. *Revista Española de Investigaciones Sociológicas*, 149, pp. 31-44. <https://doi.org/10.5477/cis/reis.149.31>

Gaskin, J. (2016). Confirmatory Factor Analysis. Recuperado de <http://statwiki.gaskination.com/index.php?title=CFA>

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57

Gaskin, J. y Lim, J. (2016). Model Fit Measures, AMOS Plugin. Recuperado de <https://drive.google.com/drive/folders/0B3TITGdHG9aEbFgleEpcOWtrR3c>

Gaskin, J. (2016). Stat Wiki. Recuperado de <http://statwiki.gaskination.com/index.php?title=CFA>

Hair J., Joseph, Black, William, Babin, Barry y Anderson, Rolph. (2014). *Multivariate Data Analysis: Pearson new international edition*. 7a Ed. Essex: Pearson. New Jersey.

Hamdan González, N. (1985). *Métodos estadísticos en educación*. 2da ed. Publicaciones Bourgeón C. A. Caracas Venezuela.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>

Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2014). *Metodología de la investigación*. 6ª ed. México: Mc Graw Hill. Recuperado de [https://periodicooficial.jalisco.gob.mx/sites/periodicooficial.jalisco.gob.mx/files/metodologia\\_de\\_la\\_investigacion\\_-\\_roberto\\_hernandez\\_sampieri.pdf](https://periodicooficial.jalisco.gob.mx/sites/periodicooficial.jalisco.gob.mx/files/metodologia_de_la_investigacion_-_roberto_hernandez_sampieri.pdf)

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. 3a ed. Nueva York: Guilford Press.

Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A. y Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de psicología*, 30(3); pp. 1151–1169. <https://dx.doi.org/10.6018/analesps.30.3.199361>

Mousalli-Kayat, G. (2015). *Métodos y diseños de investigación cuantitativa*. Universidad de los Andes, Mérida. DOI: [10.13140/RG.2.1.2633.9446](https://doi.org/10.13140/RG.2.1.2633.9446)

Muiños, J. y Federico, M. (2021). Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *PSOCIAL*, 7(1); pp. 66-71. Recuperado de <https://www.redalyc.org/articulo.oa?id=672371335005>

Nuviala Nuviala, A., García Fernández, J., Bernal-García, A., Grao-Cruces, A., Tevas-Villén, M. R. y Pérez Ordaz, R. (2014). Adaptación y validación de la escala de intenciones futuras de comportamiento en usuarios de servicios deportivos. *Universitas Psychologica* 13(3), pp.15- 25. Recuperado de <https://www.redalyc.org/pdf/647/64733438022.pdf>

Parra Velasco, L. Y. y Vásquez Martínez, M. G. (2017). Muestreo probabilístico y no probabilístico. *Universidad del Istmo*. Recuperado de <https://www.gestiopolis.com/wp-content/uploads/2017/02/muestreo-probabilistico-no-probabilistico-guadalupe.pdf>

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57

Porras Cerron, J. C. (2016). Comparación de pruebas de normalidad multivariada. *Anales Científicos* 77(2), pp. 141-146. <http://dx.doi.org/10.21704/ac.v77i2.483>

Roco Videla, A., Hernández Orellana, M. y Silva González, O. (2021). ¿Cuál es el tamaño muestral adecuado para validar un cuestionario? *Cartas al Editor*, 38(4). pp. 877-878. <https://dx.doi.org/10.20960/nh.03633>

Sánchez Carrión, J. J. (1999). Manual de análisis estadístico de los datos. 2da Ed. Alianza Editorial. S. A. Madrid España.

Salessi, S. y Omar, A. (2018). Pasión por el trabajo: evidencias de validez discriminante, predictiva e incremental. *Psicología: Ciência e Profissão* 38(3), 522-536. <https://doi.org/10.1590/1982-37030004752017>

Serramona, J. (1980). Investigación y estadística aplicadas a la educación. Ediciones CEAC. Barcelona – España.

Schumacker, Randall y Lomax, Richard. (2010). A beginner's guide to structural equation modeling (3ª ed.). Lawrence Erlbaum Associates.

Talledo Flores, H. y Sánchez Herrera, J. (2007). Del análisis conjunto a las escalas de máximas diferencias: un estudio comparativo entre MAXDIFF y escalas Monádicas, desde la validez convergente y discriminante. XIX Encuentro de profesores universitarios de marketing, 19, 20 y 21 de septiembre de 2007. Universidad Complutense de Madrid. Recuperado de [https://scholar.google.es/scholar?hl=es&as\\_sdt=0%2C5&q=Del+an%C3%A1lisis+conjunto+a+las+escalas+de+m%C3%A1ximas+diferencias%3A+un+estudio+comparativo+entre+MAXDIFF+y+escalas+Mon%C3%A1dicas%2C+desde+la+validez+convergente+y+discriminante&btnG=](https://scholar.google.es/scholar?hl=es&as_sdt=0%2C5&q=Del+an%C3%A1lisis+conjunto+a+las+escalas+de+m%C3%A1ximas+diferencias%3A+un+estudio+comparativo+entre+MAXDIFF+y+escalas+Mon%C3%A1dicas%2C+desde+la+validez+convergente+y+discriminante&btnG=)

Vargas Halabí, T. y Mora-Esquivel, R. (2017). Tamaño de la muestra en modelos de ecuaciones estructurales con constructos latentes: Un método práctico. *Revista Electrónica Actualidades Investigativas en Educación*, 17(1); pp. 1-34. <http://dx.doi.org/10.15517/aie.v17i1.27294>

Weinberg, S. L. y Goldberg, K. P. (1982) Estadística básica para las ciencias sociales. Interamericana S. A. México, D. F.

Jesús Eduardo Pulido//Tamaños de muestra y su incidencia en la invarianza y propiedades psicométricas...29-57

### Conflicto de interés

El autor de este manuscrito declara no tener ningún conflicto de interés.

### Declaración ética

El autor declara que el proceso de investigación que dio lugar al presente manuscrito se desarrolló siguiendo criterios éticos, por lo que fueron empleadas en forma racional y profesional las herramientas tecnológicas asociadas a la generación del conocimiento.

### Copyright

La *Revista de la Universidad del Zulia* declara que reconoce los derechos de los autores de los trabajos originales que en ella se publican; dichos trabajos son propiedad intelectual de sus autores. Los autores preservan sus derechos de autoría y comparten sin propósitos comerciales, según la licencia adoptada por la revista.

### Licencia Creative Commons

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-Compartir Igual 4.0 Internacional



REVISTA DE LA UNIVERSIDAD DEL ZULIA, Fundada el 31 de mayo de 1947

UNIVERSIDAD DEL ZULIA, Fundada el 11 de septiembre de 1891