

González-Montesinos, Manuel-Jorge & Backhoff, Eduardo (2010). Validación de un cuestionario de contexto para evaluar sistemas educativos con Modelos de Ecuaciones Estructurales. *RELIEVE*, v. 16, n. 2, p. 1-17.
http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_1.htm

**e-Journal of Educational
Research, Assessment and
Evaluation**



**Revista ELectrónica de
Investigación y EValuación
Educativa**

VALIDACIÓN DE UN CUESTIONARIO DE CONTEXTO PARA EVALUAR SISTEMAS EDUCATIVOS CON MODELOS DE ECUACIONES ESTRUCTURALES

[Validation of a context questionnaire for the evaluation of educational systems with Structural Equations Modeling]

by / por

[Article record](#)

[About authors](#)

[HTML format](#)

González-Montesinos, Manuel-Jorge (mgm@caborca.uson.mx)

Backhoff, Eduardo (backhoff@uabc.edu.mx)

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

Abstract

This work shows the use of Structural Equations Modeling (SEM) methodology to validate the conceptual structure of context questionnaires employed in national and international studies that evaluate the educational achievement of students. Unfortunately, evaluation agencies do not assign the same amount of technical resources to context questionnaires as those allotted to achievement tests, consequently the former are rarely subject to validity studies. In order to show the use of SEM methodology, a questionnaire was selected, which was employed by the National Institute for the Evaluation of Education to investigate the background factors associated with educational achievement of elementary school students in the third grade. It was administered to a national representative sample of 55312 students. This questionnaire is analyzed according to SEM procedures of confirmatory factor analysis (CFA). The obtained results provide evidence for the constructive validity of this tool.

Keywords

Structural Equations Modeling, Confirmatory Factor Analysis, Construct Validity, Context Questionnaires, National Institute for the Evaluation of Education

Resumen

Este trabajo tiene el propósito de mostrar el uso de la metodología de Modelos de Ecuaciones Estructurales (SEM) para validar la estructura conceptual de cuestionarios de contexto que se utilizan en los estudios nacionales e internacionales para evaluar el logro educativo de los escolares. Desgraciadamente, los organismos evaluadores no destinan los mismos recursos técnicos a los cuestionarios de contexto que a las pruebas de aprendizaje, por lo que rara vez los primeros son sujetos a estudios de validez. Para mostrar el uso de la metodología SEM, se seleccionó el cuestionario dirigido a alumnos que utilizó el Instituto Nacional para la Evaluación de la Educación (INEE) para conocer los factores de contexto asociados el logro educativo de los estudiantes de tercero de primaria, el que se aplicó a una muestra representativa nacional de 55312 alumnos. A dicho instrumento se le aplican los supuestos y procedimientos del análisis factorial confirmatorio (AFC) y se describen los resultados obtenidos, los cuales aportan evidencia de validez de constructo para el instrumento.

Descriptores

Modelos de Ecuaciones Estructurales, Análisis Factorial Confirmatorio, Validez de Constructo, Cuestionarios de contexto, Instituto Nacional para la Evaluación de la Educación.

En las últimas dos décadas hemos presenciado esfuerzos importantes de los países y de organismos internacionales para conocer la calidad de los sistemas educativos. Por ello, se han desarrollado pruebas de logro educativo con las cuales se han evaluado los resultados de aprendizaje que logran los estudiantes de ciertas edades o grados escolares, en los dominios curriculares de mayor importancia como las matemáticas y la lengua. Algunos ejemplos de estos estudios donde ha participado la población estudiantil de México son los realizados por el Instituto Nacional para la Evaluación de la Educación (INEE), la Organización para la Cooperación y el Desarrollo Económico (OCDE), la Asociación Internacional para la Evaluación del Logro Educativo (IEA, por sus siglas en inglés) y el Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE).

Sin embargo, conocer los resultados de aprendizaje de un país no es suficiente para tomar decisiones de política educativa, ya que es necesario también poder explicar, o al menos tener una idea, de las razones por las cuales se distribuye el aprendizaje de una forma u otra entre los distintos grupos de estudiantes al interior de un país, o bien entre los distintos países evaluados.

Por ello, es indispensable no sólo evaluar el aprendizaje de los estudiantes, sino también conocer las condiciones contextuales en que éste ocurre. Está muy bien documentado que las condiciones socioeconómicas de las familias de los estudiantes y las condiciones escolares donde ellos estudian están fuertemente vinculadas con los resultados educativos (Willms, 2006). Así, por lo general, los estudiantes con mejores condiciones socioeconómicas y escolares son quienes obtienen los resultados más altos en las pruebas de logro educativo en cualquier país del mundo.

Para conocer las condiciones sociales y escolares donde se desenvuelven y aprenden los estudiantes, por lo general se utilizan

cuestionarios dirigidos a estudiantes, padres de familia, profesores y directores de escuelas, cuyo propósito es obtener información de variables contextuales que se asocien con el aprendizaje y ayuden a explicar los resultados de las evaluaciones nacionales e internacionales antes referidas (OCDE, 2006).

En resumen, para evaluar la calidad de los sistemas educativos se utilizan dos tipos de instrumentos: las pruebas de aprendizaje y los cuestionarios de contexto. Es importante señalar que el proceso para diseñar, construir y validar ambos tipos de instrumentos de medición en teoría es el mismo; en los dos casos se busca que las interpretaciones basadas en la información recabada reflejen la realidad que se intenta capturar. Sin embargo, en la práctica las pruebas de aprendizaje reciben mayor atención presupuestal y técnica en todos los aspectos que los cuestionarios de contexto, a los que, por lo general, se exige menos rigor técnico y rara vez se someten a procesos de validación. En consecuencia, muchos de los esfuerzos nacionales e internacionales que intentan explicar el aprendizaje de los estudiantes no logran su propósito final, ya que los cuestionarios utilizados aportan resultados débiles y poco confiables.

Afortunadamente, en los últimos años se han venido desarrollando y afinando nuevas metodologías para validar distintos instrumentos de medición y con ello asegurar la calidad de los resultados e interpretaciones que de éstos se deriven. Este es el caso de los Modelos de Ecuaciones Estructurales (SEM, por sus siglas en inglés), los cuales se han utilizado en varios de los estudios publicados por el INEE para validar los cuestionarios de contexto que se utilizan junto con los Exámenes para la Calidad y el Logro Educativos [Excale] (Backhoff, Bouzas, González-Montesinos, Andrade, Hernández y Contreras, 2008).

Dado que estos resultados se ha difundido básicamente en los informes técnicos del INEE y no han tenido una mayor difusión en

revistas de investigación especializadas de habla hispana, este artículo tiene el propósito de mostrar el uso de la metodología SEM para validar el constructo de uno de los cuestionarios de contexto utilizado en el estudio nacional realizado por el INEE *El aprendizaje en tercero de primaria en México* (Backhoff, Andrade, Peón y Sánchez, 2007).

Para lograr este propósito, el trabajo se divide en cuatro apartados. En el primero, se hace una descripción breve de los principios de la metodología SEM. En el segundo, se describe la estructura del cuestionario de estudiantes a ser validado. En el tercero, se aplica la metodología SEM para validar el cuestionario y en el cuarto se discuten los resultados encontrados.

Modelo de Ecuaciones Estructurales

El Modelamiento de Clases Latentes (LTM por sus siglas en inglés) es una clase de análisis inferencial en el que se utiliza la regresión multivariada para relacionar patrones de respuestas a un conjunto de factores latentes que no se observan directamente pero que, según la teoría sustantiva, existen en dimensiones continuas en las personas que se evalúan (Rizopoulos, 2006).

Los supuestos que subyacen a este tipo de análisis son los siguientes:

- Un conjunto de factores latentes –por ejemplo: aptitudes, actitudes o percepciones– influye sobre un grupo de variables observadas, que se miden a través de los reactivos que componen una escala. Este supuesto se denomina *independencia condicional* y postula que las respuestas a los reactivos son independientes entre sí, pero condicionadas por la variable latente que las determina.
- Los factores latentes se pueden cuantificar por medio de una estructura de dimensiones, que se basa en una teoría sustantiva que postula la existencia de constructos psicológicos que ejercen influencia causal

sobre las respuestas de las personas a un grupo reactivos.

La versión más desarrollada del LTM es el Modelamiento de Ecuaciones Estructurales (SEM, por sus siglas en inglés), que proporciona los procedimientos y criterios técnicos más sólidos para la validación de modelos de medición bajo estos dos supuestos. La metodología SEM consta de las siguientes fases:

- 1) La especificación del modelo de medición, que establece los rasgos latentes y las dimensiones que los representan como variables de interés de una teoría sustantiva.
- 2) La implementación computarizada del sistema de ecuaciones estructurales, que se emplea para generar la evidencia de validez del modelo de medición y sus dimensiones.
- 3) El uso de índices y criterios de *bondad de ajuste* para relacionar la evidencia validatoria con la estructura dimensional del instrumento que se evalúa.
- 4) La re-especificación del modelo de medición para mejorar el ajuste del modelo añadiendo o eliminando relaciones entre los factores, cuando existe justificación teórica para ello.

En todas estas fases se utiliza una secuencia detallada de procedimientos descriptivos e inferenciales que, para efectos prácticos, puede denominarse *lógica SEM*.

Estructura del cuestionario de estudiantes de 3º de primaria

Como ya se mencionó en la introducción, anualmente el INEE realiza estudios evaluativos para conocer los niveles de logro que alcanzan los estudiantes en México. Para poder explicar la distribución de los aprendizajes y las diferencias entre los grupos de estudiantes evaluados, el INEE administra cuestionarios de contexto dirigidos a alumnos, docentes y directores para conocer las variables familiares, sociales y escolares que se asocian con el logro educativo.

Con el propósito de ejemplificar la forma de validar un instrumento de medición con la metodología SEM, en este trabajo se seleccionó el cuestionario de contexto de alumnos

de tercero de primaria, que se aplicó en 2006 a una muestra nacional de 55312 escolares de 3167 centros educativos públicos y privados (Backhoff y col., 2008).

Tabla 1. Variables que se exploran en el cuestionario de los alumnos de 3º de primaria

Nombre de la variable	Definición
VARIABLES INDIVIDUALES	
Labores domésticas	Actividades con las que el alumno contribuye al cuidado del hogar.
Edad	Edad del estudiante con relación al grado escolar. Se clasifica en normativa y extra-edad.
Sexo	Género del estudiante.
Trabajo remunerado	Actividades consideradas como trabajo asalariado.
Años en preescolar	Número de años escolares cursados antes de la primaria.
Reprobación	Condición de haber reprobado o repetido al menos un grado.
Hacinamiento	Relación del número de cuartos entre número de personas que habitan en una casa.
Limitaciones físicas	Condiciones orgánicas que limitan el aprendizaje.
Leer inglés	Auto reporte sobre el grado en que el estudiante considera que entiende el inglés escrito.
Uso de la computadora	Auto reporte sobre el grado en que el estudiante considera que conoce el uso de una computadora.
Años en escuela	Número de años que el alumno ha permanecido en la misma escuela.
Ingreso a primaria	Edad en que el alumno ingresó a primero de primaria.
Estructura familiar	Condición de vivir con ambos o alguno de los padres.
VARIABLES COMPUESTAS O ESCALAS	
Capital cultural escolar	Escala conformada con 5 ítems: la escolaridad de los padres, las expectativas educativas de los padres hacia el alumno, el número de libros en casa y el número de ocasiones en que el alumno fue al cine en un año.
Estatus socioeconómico	Escala conformada con 8 ítems: tener la Beca de Oportunidades; material del piso de la vivienda y poseer los siguientes bienes y servicios en casa: teléfono, drenaje, automóvil, microondas, computadora e Internet.
Conflictos familiares	Escala conformada con 3 ítems: frecuencia de pleitos entre el alumno y sus papás, clima general de convivencia en la familia y frecuencia de golpes físicos en los conflictos familiares.
Supervisión padres	Escala conformada con 4 ítems: supervisión de las calificaciones, aprovisionamiento de materiales escolares solicitados por la escuela, supervisión del estudio y preparación de exámenes.
Tareas	Escala conformada con 3 ítems: frecuencia con que se realizan ejercicios escolares, número de tareas elaboradas y días a la semana dedicadas al estudio en casa.
Lengua indígena	Escala conformada con 3 ítems: lengua materna, lengua que se habla en casa y lengua que se habla en la escuela.
Conductas violentas	Escala conformada con 4 ítems: haber sido castigado por mala conducta, haber participación en peleas, haber dañado el mobiliario escolar y haber amenazado a otros estudiantes.

Fuente: Backhoff y col., 2008.

El proceso que inicia con el diseño del cuestionario y termina con el análisis de la información que proveen es bastante complejo. Una etapa importante de este proceso tiene que ver con la validación empírica del modelo conceptual del cual partió, que es el foco central de este trabajo. Para validar el modelo conceptual, primero hay que conocer la estructura del cuestionario, así como las variables y dimensiones que lo conforman.

El cuestionario de alumnos se conformó con una variedad de reactivos diseñados para evaluar 13 variables individuales y 7 variables compuestas (o escalas) que se muestran en la Tabla 1. Las variables individuales son las que se exploran con un solo reactivo, mientras que las variables compuestas se evalúan con tres o más reactivos, cuyos resultados se complementan para formar una escala. En el caso del INEE, las escalas se

calibran con el modelo de Rasch (Linacre, 2005), para lo cual se centra la media en cero unidades y su desviación estándar en una.

Todas las variables fueron analizadas empíricamente, utilizando la base de datos que publica el INEE en su página de Internet. En general, estas variables muestran tener buenos indicadores psicométricos, estar relacionadas con el aprendizaje de los estudiantes y no tener un porcentaje de respuestas omitidas mayor a 3.5%. Por su parte, las escalas se comportaron dentro de lo esperado: con buenos indicadores de ajuste y una alta consistencia interna. Para mayor información sobre la construcción y validación de estas escalas consulte a Backhoff y col. (2008).

Validación del cuestionario de alumnos con SEM

Las variables y escalas descritas en el apartado anterior parten de un modelo conceptual previamente definido por el INEE, que sirvió de base para elaborar los distintos cuestionarios de contexto que se utilizan para recabar información de los estudiantes. Aunque dichas variables en lo individual tengan buenas propiedades psicométricas, es necesario mostrar evidencia de validez de constructo del modelo conceptual que las subyace, lo que es el propósito central de este apartado.

Como ya se mencionó previamente, la metodología SEM consta de cuatro fases o eta-

pas. A continuación se describen los resultados obtenidos en cada una de ellas.

Etapa I: Especificación del Modelo de Medición

Esta fase es de carácter conceptual y en ella se formula la *estructura de relaciones* entre las variables latentes representadas por las dimensiones del instrumento y las respuestas a los reactivos del cuestionario de contexto.

El Modelo conceptual de los cuestionarios de alumnos de tercero de primaria está compuesto por 20 variables observadas que se agrupan en cuatro dimensiones, tal y como se muestra en la Tabla 2. Sin embargo, es importante mencionar que las variables de sexo y edad, aunque forman parte importante de la información personal del cuestionario, no se consideraron en el modelo a evaluar, ya que en principio no se puede esperar que se relacionen con las demás variables.

En este modelo los factores latentes son las dimensiones propuestas en la Tabla 2, y las variables observadas son los valores que se obtienen de las respuestas a los reactivos del cuestionario. Una vez establecida la estructura del instrumento a nivel conceptual, ésta se postula como la teoría cuya dimensionalidad será validada, con base en los resultados obtenidos.

Tabla 2. Modelo conceptual de las dimensiones y variables de contexto: estudiantes de 3º de primaria

Dimensiones	Variables Observadas	Tipo de variable
Características Personales del Alumno (CPERS)	Sexo (SEXO)*	Individual
	Edad (EDAD)*	Individual
	Lengua indígena (LENG)	Escala
	Limitaciones físicas (LIMI)	Individual
	Conductas violentas (COVI)	Escala
Situación Sociocultural del Alumno (SOCC)	Capital cultural escolar (CCES)	Escala
	Estatus socio económico (SES)	Escala
	Hacinamiento (HACI)	Individual
	Trabajo remunerado (TRAB)	Individual
Dinámica Familiar (DFAMI)	Estructura familiar (EST)	Individual
	Conflictos familiares (CONF)	Escala
	Supervisión de los padres (SUPP)	Escala
	Labores domésticas (LABD)	Individual
Antecedentes Escolares del Alumno (AESC)	Años en preescolar (PRES)	Individual
	Ingreso a la primaria (INGR)	Individual
	Años en la escuela (NESC)	Individual
	Reprobación (RERP)	Individual
	Tareas (TARE)	Escala
	Leer inglés (ING)	Individual
	Uso de computadora (COMP)	Individual

* Variables no consideradas en el análisis SEM

Etapa II: Implementación Computacional del Sistema de Ecuaciones Estructurales

En esta segunda fase se emplea la metodología de modelación de ecuaciones estructurales SEM, implementada en el paquete LISREL 8.8, en la modalidad de Análisis Factorial Confirmatorio [AFC] (Jöreskog & Sörbom, 1996-2001).

La lógica de este procedimiento confirmatorio consiste en analizar la estructura de covarianza en la base de datos que contiene las 18 variables observadas y tratar de extraer evidencia de validez para afirmar que el modelo de medición coincide con la estructura conceptual postulada en la teoría de dimensionalidad del instrumento. Esto se logra al comparar dos estructuras de covarianza. La primera se denomina *la matriz de covarianzas derivada* de las variables observadas. La segunda se denomina *la matriz de covarianzas reproducida* por el modelo.

La noción de modelación de las relaciones estructurales se explica porque la *matriz reproducida* modela la estructura de relaciones especificada conceptualmente. Es decir, el proceso de modelación reproduce fielmente las relaciones entre las variables latentes y observadas que se

declara como hipótesis en el modelo conceptual. En esencia la comparación de las matrices se realiza al someter a prueba la siguiente hipótesis nula:

$$H_0: \Sigma = \Sigma(\theta)$$

Donde Σ es la *matriz de covarianzas derivada* de las variables observadas, y $\Sigma(\theta)$ es la *matriz de covarianzas reproducida* por el modelo hipotético de medición.

En la práctica, el procedimiento de prueba se representa con mayor exactitud bajo la siguiente hipótesis: $H_0: \Sigma \approx \Sigma(\theta)$, lo que implica que bajo la hipótesis nula, la matriz derivada aproxima a la matriz reproducida. La noción de aproximación entre estas estructuras es crucial ya que, como se verá más adelante, en realidad estas matrices pueden diferir ligeramente, dentro de límites de tolerancia rigurosamente definidos.

Es muy importante notar que en todo el procedimiento el objetivo fundamental es demostrar que las dos estructuras de covarianza no difieren significativamente entre sí. Es decir, cuando el modelo de medición y la estructura de relaciones en los datos observados ajustan entre sí, las matrices correspondientes se aproximan

y, como la diferencia entre ellas es mínima, la hipótesis nula no debe rechazarse sino retenerse. De hecho, cuando éste es el caso, el valor p del estadígrafo de prueba deber resultar mayor a 0.05, indicando así que la diferencia no es estadísticamente significativa. Como se verá más adelante, esta condición de ajuste preliminar no siempre se obtiene debido a algunas limitaciones de cálculo de este índice.

La fase de especificación del modelo incluye una representación gráfica de la estructura teórico-conceptual del instrumento bajo análisis. Esta representación es la base para la formulación de la matriz reproducida que será comparada con la matriz derivada. Para el cuestionario de contexto de los estudiantes mexicanos de tercero de primaria la representación gráfica se muestra en la figura 1.

En esta figura, los óvalos representan las cuatro variables latentes (constructos) y los recuadros las 18 variables observadas, que en conjun-

to conforman la dimensionalidad del instrumento. Las flechas bidireccionales representan covarianzas entre las variables latentes. Las flechas unidireccionales representan las líneas de influencia de las variables latentes sobre las respectivas variables observadas. Estas especificaciones de direccionalidad equivalen a afirmar que cada una de las variables relacionadas mide (o es un indicador de) la dimensión a la que en teoría pertenecen. Finalmente, las flechas unidireccionales que aparecen a la izquierda de los recuadros, representan los términos de error que son la influencia de fuentes de variabilidad única (o idiosincrática) asociadas a cada variable observada.

Para someter a prueba esta especificación de modelo se empleó el programa LISREL 8.8, cuya programación se detalla en la guía de usuario (Jöreskog & Sörbom, 1996-2001)^[1]. Siguiendo la especificación natural del modelo, la implementación del AFC produce los resultados que se muestran en la Figura 1.

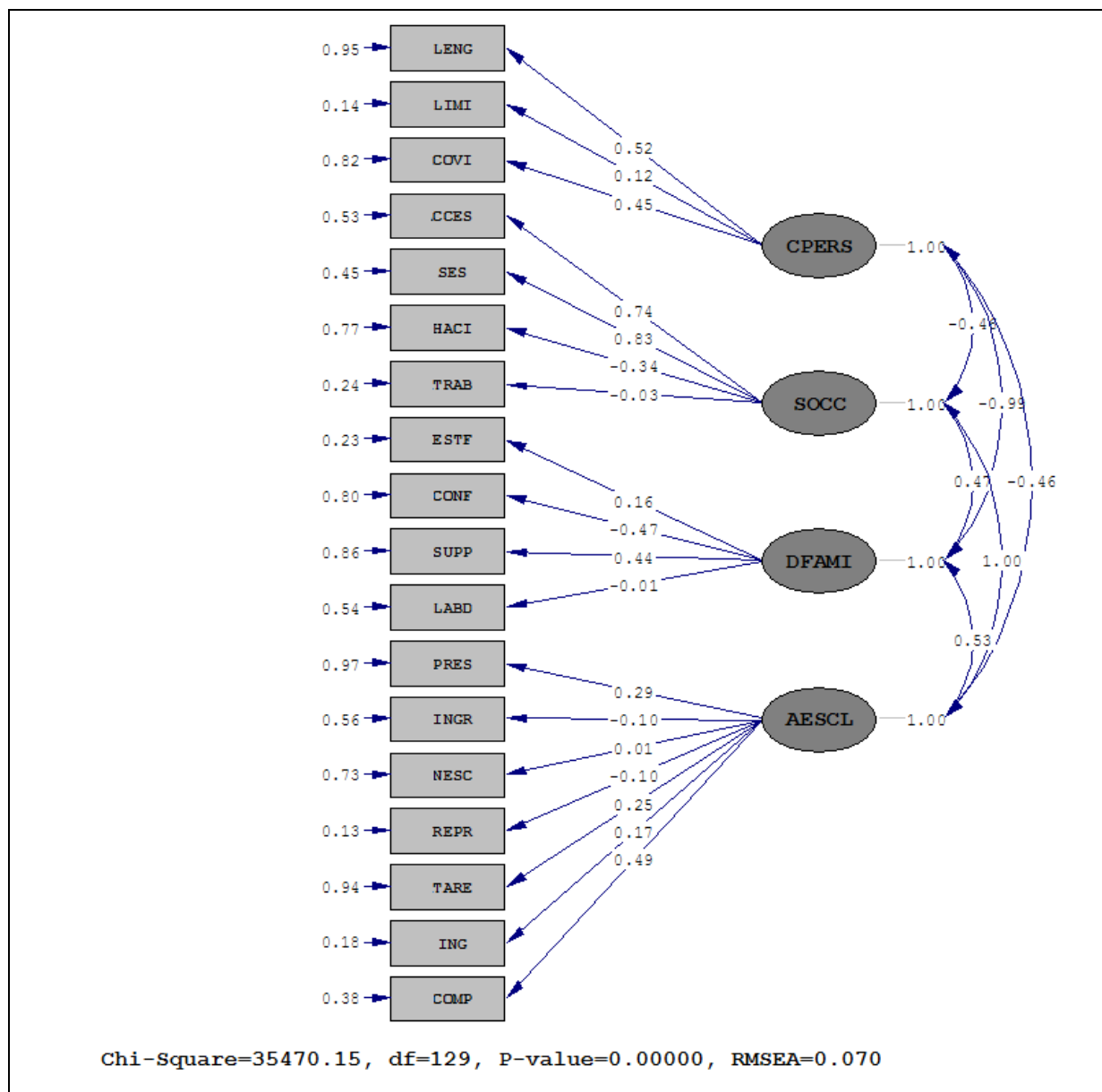


Figura 1. Representación gráfica del modelo de medición natural del cuestionario de alumnos.

En primer término, el análisis produce una representación de las relaciones estructurales observadas en la matriz derivada. Los valores que presentan las flechas unidireccionales son las líneas de influencia entre las variables latentes y observadas y se interpretan como coeficientes de regresión multivariada. Estos han sido calculados por el programa al analizar la matriz de covarianza derivada de las variables observadas. Los valores que aparecen al final de la figura son índices preliminares de bondad de ajuste para una evaluación inicial del modelo. Su aplicación e interpretación se describe más adelante.

En cuanto a los coeficientes de regresión entre las variables latentes y las observadas, la interpretación es como sigue. Los primeros tres valores son: 0.52, 0.12 y 0.45 y corresponden a la influencia de la variable latente *Características personales* (CPERS) sobre las variables observadas *Lengua indígena* (LENG), *Limitaciones físicas* (LIMI) y *Conductas violentas* (COVI). La influencia que ejerce la variable latente sobre las variables observadas indica que cuando CPERS aumenta una unidad, LENG, LIMI y COVI aumentan en proporciones de 0.52, 0.12 y 0.45, respectivamente.

Cuando los coeficientes de regresión tienen signo positivo, la relación entre la variable latente y la observada es directa. Es decir, cuando la variable latente aumenta una unidad, la variable observada también aumenta según el peso del coeficiente. Por el contrario, cuando los coeficientes de regresión tienen signo negativo, la relación entre la latente y la observada es inversa. Es decir, cuando la variable latente aumenta una unidad, la variable observada disminuye según el peso del coeficiente. Este es el caso de la relación entre la latente *Situación sociocultural* (SOCC) y las observadas *Hacinamiento* (HACI) y *Trabajo remunerado* (TRAB), donde los valores de -0.34 y -0.03 significan que cuando SOCC aumenta en una unidad, HACI y TRAB disminuyen en esa proporción. Esto indica que los alumnos con mejor situación sociocultural viven en hogares con menor hacinamiento y realizan menos trabajos remunerados. La misma situación se observa en la relación de *Dinámica familiar* (DFAMI) y *Conflictos familiares* (CONF) donde el coeficiente es -0.47. El mismo caso se observa para la variable *Labores domésticas* (LAB) con una reducción de -0.01. Esto implica que los alumnos con mejor situación en su dinámica familiar tienden a tener menos conflictos familiares y menos participación en labores domésticas.

Las relaciones entre los *Antecedentes escolares* (ESC) y las variables observadas se sintetizan a continuación. El *Año de ingreso a la escuela* (INGR) y la *Reprobación* (REPR) presentan una disminución de -0.10 respectivamente. Es importante notar que la dirección de estas relaciones tiene sentido según la teoría sustantiva, ya que es de esperar que a medida que aumentan los antecedentes escolares, el año de ingreso a la escuela y la reprobación tienden a disminuir. Es decir, los alumnos con mejores antecedentes escolares tienden a ingresar a la escuela más temprano y tienden a reprobarnos menos.

Redondeando esta explicación, se observa que los coeficientes positivos en la dimensión de *Antecedentes escolares* (ESC) son: 0.29 para *Años en Preescolar* (PRE), 0.01 para *Años de escolaridad* (NESC), 0.25 para *Tareas* (TARE), 0.17 para *Uso del inglés* (ING) y 0.49 para *Uso de computadora* (COMP), significan que todas estas variables tienden a aumentar en alumnos con mejores antecedentes escolares.

El mismo tipo de interpretación se debe realizar con las relaciones entre la variable latente *Situación socio cultural* (SOCC) y las variables observadas *Capital cultural escolar* (CCES), con un coeficiente de 0.74, y con *Estatus socioeconómico* (SES), con un coeficiente de 0.83.

En resumen, los resultados de esta etapa muestran que la evaluación preliminar del modelo verifica que los coeficientes derivados hayan tenido el sentido positivo y negativo que se esperaría de acuerdo con la teoría sustantiva empleada para configurar el modelo de medida.

Etapa III: Aplicación de índices y criterios de bondad de Ajuste

La tercera etapa del proceso consiste en evaluar la bondad de ajuste entre la *matriz derivada* de los datos y la *matriz reproducida* por el modelo. Como se expresó en la sección anterior, se somete a prueba la hipótesis. Se espera que la diferencia entre las dos matrices no sea estadísticamente significativa, por lo que se puede afirmar que ambas matrices se aproximan, y por lo tanto que el modelo de medición y los datos observados ajustan entre sí.

A través del desarrollo de la metodología SEM se han propuesto una serie de índices y criterios de bondad de ajuste para determinar si la condición de aproximación entre las matrices se sostiene. Si bien los índices de bondad de ajuste son muy diversos, todos tienen en común la siguiente característica: *son una medida de la diferencia absoluta*

observada entre cada uno de los elementos de la matriz derivada y la matriz reproducida. Cuando las diferencias detectadas entre los elementos son mínimas, se concluye que las estructuras de ambas matrices se aproximan y por lo tanto se retiene la hipótesis nula, concluyendo que existe un ajuste entre el modelo conceptual y las respuestas de los estudiantes.

El primer índice de bondad de ajuste es el estadígrafo de χ^2 (Chi cuadrada) que se propone para representar el punto de mejor ajuste o discrepancia mínima entre las matrices comparadas (Blalock, 1964). Con este índice se evalúa el ajuste general del modelo de medición. El valor de χ^2 representa una medida preliminar de la diferencia registrada por el procedimiento entre el modelo de medición y la estructura de covarianza observada en los datos. Para este valor, se espera obtener un mínimo de diferencia, que no resulte estadísticamente significativo. Esto implica que no hay diferencia apreciable entre el modelo propuesto y los datos empíricos. Debe notarse que el resultado de esta prueba es preliminar y si la diferencia resulta estadísticamente significativa, el resultado se toma como base para el cálculo de otros índices de ajuste. También es importante asentar que este valor, aún cuando no se interpreta directamente cuando su valor p es menor a 0.05, se retiene para usos posteriores, puesto que es de utilidad en pruebas comparativas entre modelos.

Como se puede observar en la Figura 1, el valor de χ^2 resultó igual a 35470.15 ($p=0.000$). Debe notarse que el valor de p es menor a 0.05 lo que indica que la diferencia entre las matrices de covarianzas comparadas (la derivada y la reproducida) es estadísticamente significativa. Por sí solo, este resultado implica que el modelo y los datos no ajustan entre sí. Sin embargo, este resultado preliminar no es definitivo, ya que este estadígrafo tiene la limitación de ser muy sensible a mínimas diferencias cuando el tamaño de

la muestra es muy grande. En este caso, como se recordará fue de 55312 escolares. Cuando se presenta esta situación, se omite la interpretación inicial de χ^2 y se procede a evaluar el ajuste del modelo con índices de bondad de ajuste que han sido desarrollados para superar la hipersensibilidad de χ^2 .

El índice de bondad de ajuste más robusto propuesto a la fecha es el *Error Medio Cuadrático de Aproximación* (RMSEA, por sus siglas en inglés). Este índice ha sido desarrollado como una medida absoluta de la diferencia de la estructura de relaciones entre el modelo propuesto y los valores de covarianza en población medida (Steiger, 1990).

Su cálculo es como sigue:

$$RMSEA = \sqrt{\frac{\widehat{\delta}_M}{df_M(N-1)}}$$

En este caso el término $\widehat{\delta}_M$ proviene de la medida inicial de diferencia entre los datos y el modelo:

$$\widehat{\delta}_M = \max(\chi_M^2 - df_M, 0)$$

Lo anterior significa que el valor mayor de discrepancia entre el modelo y los datos observados $\widehat{\delta}_M$, se obtiene a partir del término con mayor valor, ya sea χ^2 menos los grados de libertad del modelo, o bien cero. La importancia de este índice radica en que refleja una diferencia absoluta entre el modelo propuesto y los datos observados, tomando en cuenta el número de estimaciones y el tamaño de la muestra implicada por el modelo bajo prueba. Es muy importante notar que este índice, debido a su origen y propiedades estadísticas, compara el modelo con la estructura de relaciones entre las variables en la población.

Los lineamientos de interpretación de este índice son:

- Cuando el RMSEA presenta valores menores a 0.10 se tiene una indicación de

buen ajuste entre el modelo de medición y la estructura de los datos.

- Cuando los valores del RMSEA resultan menores a 0.05 el ajuste entre el modelo y los datos es considerado superior.
- Cuando los valores del RMSEA resultan menores a 0.01 el ajuste entre el modelo y los datos es sobresaliente.

En la Figura 1 se observa que el valor del RMSEA fue de 0.70, lo que significa que el modelo de medición del cuestionario de alumnos y la estructura de covarianzas de las respuestas de los estudiantes tienen un ajuste razonable.

No obstante lo anterior, por razones prácticas debe incorporarse evidencia que refuerce los resultados obtenidos. Para ello se recurre a interpretar otro índice de bondad de ajuste: el *Índice de Ajuste Comparativo* (CFI, por sus siglas en inglés) (Bentler, 1990). Su cálculo es como sigue:

$$CFI = 1 - \left[\frac{(\chi^2_{Model} - df_{Model})}{\chi^2_{Indep} - df_{Indep}} \right]$$

Este índice compara los valores de discrepancia de χ^2 para dos casos: la discrepancia del modelo hipotetizado, y la discrepancia de un modelo teóricamente posible (llamado modelo de independencia), en el que todas las relaciones entre variables son nulas. Es decir, este índice compara la estructura de covarianzas del modelo de medición contra una situación hipotética donde no existe relación alguna entre las variables observadas. Los valores del CFI oscilan entre 0 y 1. Cuando el CFI toma valores que exceden 0.90 se considera que existe ajuste razonable entre el modelo y los datos. A medida que el CFI aproxima 1.00 el ajuste es mejor y llega a ser sobresaliente a partir de 0.99.

El valor del CFI del modelo evaluado fue de 0.83, lo que significa que el modelo de medición del cuestionario y la estructura de los datos se aproximan a un ajuste razonable, pero no logran el mínimo aceptable de 0.90.

A partir de este resultado no es posible afirmar que la hipótesis nula se sostiene. Es decir, debe rechazarse la hipótesis establecida inicialmente $H_0: \Sigma \approx \Sigma(\theta)$. Esta situación obliga a revisar el modelo y a repetir el análisis.

Aunque en esta etapa los resultados del modelo de medición de alumnos no son favorables, para efectos de ilustración de los criterios de interpretación conjunta, se revisarán dos índices de ajuste más: el *Índice de Ajuste Incremental* (IFI, por sus siglas en inglés) y la *Raíz Cuadrática Media de Residuales* (RMSR, por sus siglas en inglés).

El cálculo de IFI es como sigue:

$$IFI = \frac{(\chi^2_{Indep} - \chi^2_{Model})}{\chi^2_{Indep} - df_{Model}}$$

El IFI es útil para comparar el ajuste relativo de dos o más modelos cuando se realizan análisis repetidos. Los resultados de este análisis muestran que el modelo de medición de alumnos tiene un IFI de 0.83. Aunque no se postule un modelo en competencia es posible interpretar este valor. Para este caso debe notarse que coincide con el CFI. Este índice toma valores entre 0 y 1, y los valores que aproximen 1 son indicadores de mejor ajuste del modelo en comparación con un modelo alternativo.

Finalmente para completar este ejercicio de interpretación, se agrega el índice RMSR, que es la raíz cuadrada de la media de las diferencias residuales detectadas al comparar cada uno de los elementos de las matrices de interés. Es decir este índice es una medida absoluta de las discrepancias entre cada uno de los elementos de la matriz de covarianzas reproducida (Γ) por el modelo y la matriz de covarianzas derivada de los datos muestrales observados (O). Su cálculo es como sigue:

$$RMSR = \sqrt{\frac{\sum(\Gamma - O)^2}{k}}$$

Para efectos prácticos se interpreta este valor en su versión estandarizada SRMR. Se considera que cuando este valor resulta menor a 0.05 existe ajuste razonable. El valor obtenido para el cuestionario de alumnos es de 0.065, valor que rebasa el límite establecido y se concluye que el modelo y los datos no ajustan.

Etapa IV: Re-especificación del modelo de medición para mejorar el ajuste

Ya que el modelo original con el que se diseñó el cuestionario de contexto de los alumnos de tercero de primaria no cumplió con los criterios de ajuste descritos anteriormente, fue necesario aplicar recursos de revisión para obtener una re-especificación de las relaciones postuladas en el modelo original y con ello mejorar el ajuste a las respuestas de los estudiantes con el modelo teórico. Los recursos de revisión empleados se derivan de la información de salida de LISREL con la opción de *Índices de Modificación* (MI). Este recurso muestra relaciones en el modelo original y permite incluirlas en la re-especificación. En concreto por cada índice de modificación sugerido se presenta la cantidad en que el valor de χ^2 (discrepancia) se reduciría al incluir la relación adicional que se detecta en la matriz de covarianzas derivada de los datos empíricos.

La Figura 2 muestra los resultados de un modelo de medición que cumple con los cri-

terios de ajuste que exige la metodología SEM. El procedimiento de re-especificación sugirió hacer los siguientes cambios: trazar una línea de influencia entre las variables *Trabajo remunerado* (TRAB) y *Labores domésticas* (LAB); relacionar el factor *Características personales* (CPERS) con la variable *Tareas* (TARE), y; asociar el factor *Dinámica familiar* (DFAMI) con la variable *Reprobación* (REPR).

En esta figura se pueden observar los resultados de un modelo de medición que cumple con los criterios de ajuste que exige la metodología SEM. Es importante señalar que las nuevas relaciones planteadas entre los elementos del modelo traen como consecuencia un cambio en los valores de los coeficientes de regresión. De inicio, puede observarse que existe una relación de covarianza de 0.05 entre los términos de error de las variables TRAB y LAB. Este valor se interpreta como que estas dos variables observadas comparten una fuente alterna de varianza distinta a las variables latentes especificadas en el modelo. También se deben destacar nuevas relaciones entre dos dimensiones y dos variables, que en el modelo original no estaban definidas: la relación negativa de la dimensión CPERS con la variable TARE, y el factor DFAMI que se relaciona negativamente con la variable REPR. En ambos casos, estas nuevas relaciones se pueden explicar tomando en cuenta las variables que componen cada factor^[2].

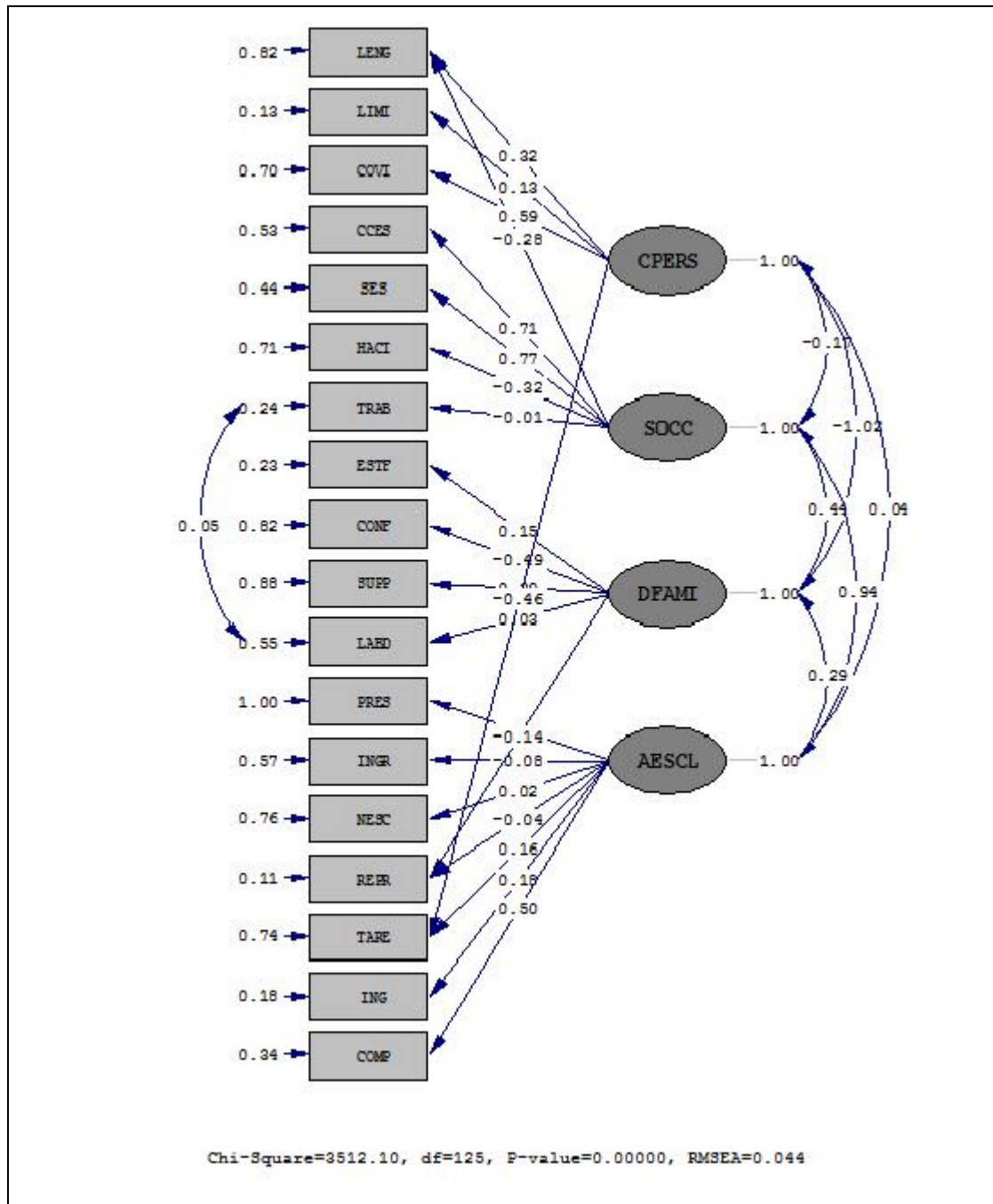


Figura 2. Representación grafica del nuevo modelo de medición del cuestionario de alumnos.

Como resultado de la re-especificación del modelo, se obtiene un índice de modificación que reduce el valor de χ^2 en 2,700.97 unidades, lo que mejora los índices globales de bondad de ajuste del modelo empírico, que se sintetizan en la Tabla 3. Como se podrá apreciar las modificaciones basadas en evidencia métrica que se incluyeron al modelo original hacen que se cumpla con cuatro de

los cinco indicadores analizados. Sólo en el indicador de χ^2 el nuevo modelo no cumplió con el criterio establecido. Pero como ya se había explicado en un inicio, este indicador es excesivamente sensible cuando se utilizan muestras poblacionales muy grandes, como fue este caso.

Tabla 3. Indicadores de ajuste del modelo de medición

Índice	Valor obtenido	Criterio límite	Interpretación
χ^2	0.00	> 0.50	No cumple
RMSEA	0.044	< 0.05	Cumple
CFI	0.92	> 0.90	Cumple
IFI	0.92	> 0.90	Cumple
SMRS	0.044	<0.05	Cumple

Discusión y conclusiones

Cada día es más común que los estudios sobre logro educativo incluyan cuestionarios de contexto que complementen los resultados de las pruebas de aprendizaje, con el propósito de ayudar a explicar las diferencias que se observan entre los distintos grupos de estudiantes evaluados. Los estudios evaluativos de la OCDE, conocidos como PISA (*Programme of International Student Assessment*), los del INEE, los del Laboratorio Latinoamericano de Calidad Educativa (LLECE) y los de la IEA, conocidos como TIMSS (*Trends in International Mathematics and Science Study*) tienen en común esta particularidad.

No obstante la importancia que tienen estos instrumentos para evaluar el contexto en que se desarrollan y aprenden los estudiantes, hoy en día persiste la idea de considerar a estos instrumentos con una importancia secundaria, por lo que se les exige poco rigor metodológico en su diseño, construcción y validación. Por esta razón, no es común que en los estudios nacionales e internacionales aporten evidencia de validez de los cuestionarios utilizados.

Tomando esto en consideración, este trabajo se propuso estudiar la validez de constructo de un cuestionario utilizado a nivel nacional por el INEE con la metodología SEM, a fin de mostrar la lógica, procedimientos e índices de esta herramienta estadística y poder hacer recomendaciones para mejorar el modelo conceptual en el cual se sustentan los cuestionarios de estudiantes utilizados por el Instituto.

Se puso el énfasis en la validez de constructo, dado que ésta se define como la propiedad métrica que asegura que los datos o valores resultantes de la aplicación de un instrumento son efectivamente interpretables como manifestación de los rasgos latentes que se miden, de acuerdo con la teoría sustantiva y en el contexto de diseño del modelo de medición correspondiente.

En síntesis, los resultados de este trabajo muestran que:

- En principio, las relaciones entre las dimensiones consideradas y las variables observadas tienen la congruencia esperada. Es decir, existen relaciones positivas y negativas entre estos dos elementos donde la teoría lo postula.
- Para evaluar el ajuste entre el modelo conceptual y el empírico se utilizaron en forma complementaria cinco índices de bondad de ajuste: χ^2 , RMSEA, CFI, IFI y SRMR. Los resultados para el modelo bajo análisis resultaron favorables, ya que se lograron obtener cuatro valores aceptables en los índices descritos, con excepción de χ^2 .
- Lo anterior pone en evidencia de manera muy clara que recurrir a un solo índice o incluso a dos de ellos no sería una estrategia apropiada para juzgar y validar el modelo.
- Dado que se confirmó la validez de constructo del modelo o estructura conceptual del cuestionario de contexto de los alumnos de tercero de primaria, se concluye que el instrumento bajo análisis cuenta con evidencia de validez de constructo a partir de los resultados que se reportan en este trabajo.

Terminamos señalando que aunque resulte evidente, es indispensable seleccionar, aplicar y mantener un esquema riguroso de control de calidad para determinar y respaldar las propiedades métricas de los instrumentos que se emplean en un proceso de investiga-

ción. Junto a un planteamiento teórico acucioso y a un diseño metodológico apropiado, un proceso de investigación requiere datos de precisión. La precisión de los datos con que se trabaja depende de las propiedades métricas de los instrumentos que se emplean.

La generación de evidencia de validez de constructo es la etapa final y definitiva en la validación de instrumentos. Cuando esta etapa se realiza con rigor, los resultados robustecen la conclusión de que el instrumento en cuestión efectivamente captura los rasgos de interés bajo las dimensiones apropiadas conforme a la teoría sustantiva que orienta su diseño. En términos prácticos, esto equivale a validar de manera parcial pero robusta, el modelo de medición que subyace a un instrumento determinado.

Referencias bibliográficas

Antillón, Luz Elena, Larrazolo, Norma y Backhoff, Eduardo (2006). Igualación equipercentil del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *RELIEVE*, v. 12, n. 2. http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_2.htm. Consultado el 30/07/2010

Backhoff, E., Andrade, E., Peón, M. y Sánchez, A. (2007). *El aprendizaje en tercero de primaria en México*. México, D.F.: Instituto nacional para la Evaluación de la Educación.

Backhoff, E., Bouzas, A., González-Montesinos, M., Andrade, E., Hernández, E. y Contreras, C. (2008). *Factores asociados al aprendizaje de estudiantes de 3º de primaria en México*. México, D.F.: Instituto nacional para la Evaluación de la Educación.

Backhoff, E., Monroy, L., Tanamachi, M.L. (2005). *Cuestionarios de contexto de las pruebas Excale: Marco de referencia*.

México, D.F.: Instituto nacional para la Evaluación de la Educación.

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.

Blalock, H.M. (1964). *Causal inference in non-experimental research*. Chapel Hill. University of North Carolina Press.

Joreskog, K. y Sorbom, D. (1996-2001). *LISREL 8: User's Reference Guide*. Illinois: Scientific Software International, Lincolnwood.

Kelloway, E.K. (1998). *Using LISREL for Structural Equations Modeling*. Thousand Oaks, CA.: Sage Publications.

Linacre, J.M. (2005). *WINSTEPS. Rasch measurement computer program*. Chicago, IL: Winsteps.com.

Organización para la Cooperación y Desarrollo Económico [OCDE] (2006). *PISA 2006 Science Competencies for Tomorrow's World*. París: autor.

Rizopoulos, D. (2006). Ltm: An R Package for Latent Trait Modeling and Item Response Theory Analyses. *Journal of Statistical Software*, Vol. 17, 5.


Steiger, J.H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.

Notas

[1] Otra fuente recomendable para consulta de procedimientos es Kelloway (1998).

[2] Por ejemplo, las variables que componen la dimensión Características personales (Lengua indígena, Limitaciones físicas y Conductas violentas) tienden a influir negativamente con el cumplimiento de las tareas de los alumnos.

ABOUT THE AUTHORS / SOBRE LOS AUTORES

González-Montesinos, Manuel-Jorge (mgm@caborca.uson.mx). Profesor-Investigador en el Departamento de Ciencias Sociales de la Universidad de Sonora, Unidad Regional Norte, Campus Caborca (México). Sus principales líneas de investigación son Técnicas de Investigación y Métodos de Análisis Social, fundamentalmente aplicados al ámbito educativo. [Buscar otros artículos de este autor en Google Académico / Find other articles by this author in Scholar Google](#) 

Eduardo Backhoff (backhoff@uabc.edu.mx). Licenciado en psicología por la Universidad Nacional Autónoma de México, Maestro en Educación por la Universidad de Washington y Doctor en Educación por la Universidad Autónoma de Aguascalientes. Su campo de interés es el desarrollo y validación de pruebas de aprendizaje de gran escala y la evaluación asistida por computadora. Ha sido Director de Pruebas y Medición del Instituto Nacional para la Evaluación de la Educación (INEE) de México. Actualmente trabaja en el Instituto de Investigación y Desarrollo Educativo, Unidad Ensenada, de la Universidad Autónoma de Baja California (México). [Buscar otros artículos de este autor en Google Académico / Find other articles by this author in Scholar Google](#)



González-Montesinos, Manuel-Jorge & Backhoff, Eduardo (2010). Validación de un cuestionario de contexto para evaluar sistemas educativos con Modelos de Ecuaciones Estructurales. *RELIEVE*, v. 16, n. 2, p. 1-17.
http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_1.htm

ARTICLE RECORD / FICHA DEL ARTÍCULO

Reference / Referencia	González-Montesinos, Manuel-Jorge & Backhoff, Eduardo (2010). Validación de un cuestionario de contexto para evaluar sistemas educativos con Modelos de Ecuaciones Estructurales. <i>RELIEVE</i> , v. 16, n. 2. http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_1.htm
Title / Título	Validación de un cuestionario de contexto para evaluar sistemas educativos con Modelos de Ecuaciones Estructurales. [<i>Validation of a context questionnaire for the evaluation of educational systems with Structural Equations Modeling</i>].
Authors / Autores	González-Montesinos, Manuel-Jorge & Backhoff, Eduardo
Review / Revista	RELIEVE (Revista ELectrónica de Investigación y EValuación Educativa / <i>E-Journal of Educational Research, Assessment and Evaluation</i>), v. 16, n. 2.
ISSN	1134-4032
Publication date / Fecha de publicación	2010 (Reception Date: 2010 September 10; Approval Date: 2010 December 21; Publication Date: 2010 December 21).
Abstract / Resumen	<p><i>This work shows the use of Structural Equations Modeling (SEM) methodology to validate the conceptual structure of context questionnaires employed in national and international studies that evaluate the educational achievement of students. Unfortunately, evaluation agencies do not assign the same amount of technical resources to context questionnaires as those allotted to achievement tests, consequently the former are rarely subject to validity studies. In order to show the use of SEM methodology, a questionnaire was selected, which was employed by the National Institute for the Evaluation of Education to investigate the background factors associated with educational achievement of elementary school students in the third grade. It was administered to a national representative sample of 55,312 students. This questionnaire is analyzed according to SEM procedures of confirmatory factor analysis (CFA). The obtained results provide evidence for the constructive validity of this tool.</i></p> <p>Este trabajo tiene el propósito de mostrar el uso de la metodología de Modelos de Ecuaciones Estructurales (SEM) para validar la estructura conceptual de cuestionarios de contexto que se utilizan en los estudios nacionales e internacionales para evaluar el logro educativo de los escolares. Desgraciadamente, los organismos evaluadores no destinan los mismos recursos técnicos a los cuestionarios de contexto que a las pruebas de aprendizaje, por lo que rara vez los primeros son sujetos a estudios de validez. Para mostrar el uso de la metodología SEM, se seleccionó el cuestionario dirigido a alumnos que utilizó el Instituto Nacional para la Evaluación de la Educación (INEE) para conocer los factores de contexto asociados el logro educativo de los estudiantes de tercero de primaria, el que se aplicó a una muestra representativa nacional de 55, 312 alumnos. A dicho instrumento se le aplican los supuestos y procedimientos del análisis factorial confirmatorio (AFC) y se describen los resultados obtenidos, los cuales aportan evidencia de validez de constructo para el instrumento.</p>
Keywords / Descriptores	<i>Structural Equations Modeling, Confirmatory Factor Analysis, Construct Validity, Context Questionnaires, National Institute for the Evaluation of Education.</i> Modelos de Ecuaciones Estructurales, Análisis Factorial Confirmatorio, Validez de Constructo, Cuestionarios de contexto, Instituto Nacional para la Evaluación de la Educación.
Institution / Institución	(1) Instituto de Investigación y Desarrollo Educativo. Universidad Autónoma de Baja California (México). (2) Departamento de Ciencias Sociales. Universidad de Sonora (México).
Publication site / Dirección	http://www.uv.es/RELIEVE
Language / Idioma	Español (Title, abstract and keywords in English)

RELIEVE

Revista ELectrónica de Investigación y EValuación Educativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).