

EVIDENCIAS DE VALIDEZ DEL CUESTIONARIO PARA DOCENTES DEL ESTUDIO INTERNACIONAL SOBRE ENSEÑANZA Y APRENDIZAJE (TALIS-2008)

Validity evidence of the teacher's questionnaire in the Teaching and Learning International Study (TALIS-2008)

EDUARDO BACKHOFF ESCUDERO
Universidad Autónoma de Baja California

MANUEL JORGE GONZÁLEZ-MONTESINOS MARTÍNEZ
Universidad de Sonora

TALIS-2008 es el primer estudio de la OCDE, cuyo propósito es conocer y comparar las condiciones escolares en que se desarrolla la enseñanza y el aprendizaje en los planteles de tercer grado de secundaria de 23 países. Se utilizaron encuestas a muestras nacionales de docentes y directores, que fueron diseñadas en inglés y francés, para después ser traducidas a distintos idiomas. Este trabajo aporta evidencias de validez de constructo de tres escalas: creencias docentes, prácticas pedagógicas y cooperación profesional de los profesores. Se utilizaron técnicas de la modelización de ecuaciones estructurales, en la modalidad de análisis factorial confirmatorio, con las cuales se analizaron las opiniones de 62.650 docentes mexicanos que participaron en el estudio ampliado TALIS. Para el caso de México, los resultados muestran que dos escalas docentes tienen niveles de ajuste aceptables, pero no es el caso para la escala de creencias docentes, cuyos niveles de ajuste no alcanzaron los niveles mínimos deseables (aun después de haberse reestructurado). Se concluye que es importante aportar evidencias de validez de constructo de las escalas que se utilizan antes de realizar los estudios a gran escala. En síntesis, la capacidad explicativa de las escalas va a ser clave para poder aportar recomendaciones de mejora y, en consecuencia, mejorar la utilidad de este tipo de evaluaciones.

Palabras clave: TALIS, Validez de cuestionarios, Secundaria, Creencias docentes, Prácticas pedagógicas, Modelización de ecuaciones estructurales, Modelización de rasgos latentes, Análisis factorial confirmatorio, Validez de constructo, México.

En el año 2008 la Organización para la Cooperación y el Desarrollo Económico (OCDE) realizó el Estudio Internacional sobre la Enseñanza y el aprendizaje (TALIS, por sus siglas en inglés), en el que participaron 23 países: 17 de ellos miembros de la OCDE, entre los que se

encuentra México. TALIS se propuso recabar información sobre los procesos de enseñanza y las actividades de gestión escolar de las escuelas secundarias de diferentes partes del mundo. Con esta intención, en cada país se encuestaron aproximadamente a 200 directores y 4.000 profesores.

La OCDE abrió la posibilidad de que los países ampliaran el tamaño de sus muestras con el fin de que los resultados tuvieran una representatividad regional o estatal, opción que adoptó México (además de Islandia). Los resultados de este estudio se publicaron en dos reportes técnicos que proporcionan información sobre aspectos básicos de los centros escolares, de sus docentes y de sus directores: un informe internacional, donde se comparan los 23 países participantes (OCDE, 2009) y un informe nacional, donde se comparan los resultados desagregados de México por modalidad educativa y entidad federativa, con excepción de Michoacán que no participó en el estudio (Backhoff, Andrade, Bouzas, Santos y Santibáñez, 2009).

Como ya se mencionó, el propósito de TALIS fue el generar información sobre las escuelas secundarias que ayude a las autoridades a tomar decisiones para mejorar los servicios educativos que se ofrecen en este nivel escolar. Uno de los focos de atención de TALIS estuvo relacionado con las características de los profesores, entre las que se encuentran: sus creencias sobre el proceso de enseñanza-aprendizaje, las prácticas pedagógicas que realizan con sus alumnos y las actividades de cooperación profesional que llevan a cabo con otros docentes.

Para investigar estas y otras características de los docentes, TALIS diseñó un cuestionario dirigido a profesores que se conformó básicamente con preguntas de selección simple y múltiple, que fue traducido y adaptado a diferentes idiomas de acuerdo con las necesidades de los países participantes. Adicionalmente, la mayoría de los análisis de TALIS requirieron elaborar diversas escalas a partir de grupos de preguntas que, en teoría, miden un mismo constructo, como se describirá más adelante.

Desgraciadamente, ninguno de los dos reportes de TALIS incluye evidencias de validez sobre las medidas utilizadas y en los tres casos de interés de este trabajo (creencias, prácticas y colaboración profesional de docentes) el reporte

internacional de TALIS reporta que dichas medidas no mostraron invariancia entre los países, razón por la cual los resultados no pudieron ser comparados en la forma tradicional (diferencia de medias) y tuvieron que ser analizados solo a nivel nacional (OCDE, 2009), con el uso de puntuaciones *ipsativas* (anglicismo de *ipsative*). La razón de la falta de invariancia cultural es desconocida, pero al menos se puede pensar en tres grandes razones que la puede explicar: 1) la falta de validez de los constructos utilizados (Messick, 1995), 2) deficiencias en los reactivos para medir los rasgos correspondientes (Buckley, 2009) y 3) problemas en las traducciones de los reactivos que atentan contra la validez de constructo de las medidas (Solano-Flores, Backhoff y Contreras Nino, 2009).

Independientemente del problema de invarianza, es importante señalar la necesidad de aportar evidencias de validez sustantiva en todos los estudios donde se utilicen instrumentos de medición, ya que como regla universal se busca que las interpretaciones basadas en la información recabada refleje la realidad que se intenta capturar (Kane, 2001). Sin embargo, en la práctica, se exige poco rigor técnico al diseño y validación de los cuestionarios de opinión y rara vez se someten a procesos de validación; en consecuencia, una gran cantidad de cuestionarios utilizados en las ciencias sociales aportan resultados débiles y poco confiables (González-Montesinos y Backhoff, 2010).

Afortunadamente, hoy en día existen nuevos métodos para aportar evidencias de validez de los diversos instrumentos de medición y, con ello, asegurar la mínima calidad necesaria para poder soportar las interpretaciones que de estos se deriven. Este es el caso del modelamiento de rasgos latentes (LTM, por sus siglas en inglés), en el que se utiliza la regresión multivariada para observar si los patrones de respuestas que emiten los sustentantes se relacionan con los rasgos o factores que se pretenden medir, pero que no se pueden observar directamente (Rizopoulos, 2006).

Dada la importancia que tienen los estudios educativos internacionales para la toma de decisiones en materia de política educativa, el propósito de este trabajo es aportar evidencias de validez de constructo de las escalas docentes del estudio nacional de TALIS, que se utilizaron para evaluar tres tipos de rasgos de los profesores de tercer grado de secundaria en México: sus creencias docentes, sus prácticas pedagógicas y el tipo de colaboración profesional que practican.

Método

Esta investigación es de carácter retrospectivo y transversal, dado que se basa en el análisis de la información generada por el estudio TALIS que se encuentra almacenada en una base de datos disponible al público¹. A continuación se describe la base de datos de México y la información seleccionada para su análisis, así como los procedimientos estadísticos utilizados para aportar evidencias de validez.

Base de datos nacional de TALIS

La base de datos (BD) mexicana de TALIS contiene información de una muestra representativa de escuelas y docentes del país, así como de 31 entidades federativas (Michoacán no participó en el estudio). La información proviene de las respuestas que emitieron 5.645 directores y 62.650 docentes a las preguntas que se les formularon en los dos cuestionarios diseñados *ad hoc* para el estudio.

En el caso de los directores, el cuestionario consistió en 37 reactivos que solicitaron 180 respuestas; en el caso de los docentes, el cuestionario se conformó con 44 preguntas que solicitaron 190 respuestas. Las preguntas del cuestionario fueron en su gran mayoría de opción simple (donde se tiene que seleccionar alguna alternativa) y selección múltiple (donde se tienen que seleccionar dos o más alternativas).

Del total de respuestas contenidas en la BD, se seleccionaron las 30 respuestas a los reactivos que se relacionan con los tres constructos analizados y que se utilizan en los dos informes de TALIS (el nacional y el internacional). Para el caso de las creencias docentes se seleccionaron ocho reactivos, ocho más para la escala de prácticas pedagógicas y, finalmente, 14 para el apartado de colaboración profesional. La descripción de estos reactivos y la forma en cómo se escalan se describe en el apartado de resultados.

Análisis estadísticos realizados

Para mostrar evidencias de validez de constructo de las tres escalas de interés en este trabajo se seleccionó el modelamiento de ecuaciones estructurales (SEM, por sus siglas en inglés), por considerar que es la versión más desarrollada del modelamiento de rasgos latentes y porque proporciona los procedimientos y criterios técnicos más sólidos para la validación de escalas de medición (Steiger, 1990). Todos los análisis se realizaron empleando máxima verosimilitud (ML) como método de estimación

Para cada una de las tres escalas, se siguieron los pasos que a continuación se describen:

- *Especificación del modelo de medición.* Establecimiento del rasgo latente y los reactivos que lo representan, que se sustentan en el marco de referencia de TALIS-2008.
- *Implementación del sistema de ecuaciones estructurales.* Procedimiento que se emplea para generar la evidencia de validez del modelo de medición de la escala.
- *Uso de índices y criterios de bondad de ajuste.* Evidencias de ajuste del modelo teórico de TALIS con el modelo empírico bajo estudio (respuestas de los profesores).
- *Re-especificación del modelo de medición.* Cambios en las relaciones de las variables que componen el modelo de medición, para mejorar los índices de ajuste del modelo original.

A partir del segundo paso se emplea la metodología SEM, implementada en el paquete LISREL 8.8, en la modalidad de análisis factorial confirmatorio (AFC) (Jöreskog y Sörbom, 1996-2001).

Resultados

A continuación se presentan los resultados para cada una de las tres escalas analizadas en este trabajo. En cada caso, se describen las variables (reactivos) que conforman cada escala y los resultados de los cuatro pasos que contempla el modelamiento SEM, descritos en el apartado anterior.

Creencias docentes acerca de la naturaleza de la enseñanza y el aprendizaje

Como ya se mencionó, la estrategia que utilizó TALIS para evaluar las creencias acerca de la naturaleza de la enseñanza y del aprendizaje fue contrastar dos enfoques pedagógicos aparentemente opuestos: la instrucción directa y la visión

constructivista. De acuerdo con sus autores (por ejemplo, Kim, 2005; Petersen, Fennema, Carpenter y Loef, 1989), la instrucción directa tiene sus orígenes en el conductismo, en el que el aprendizaje implica formar lazos de unión entre diferentes estímulos y respuestas, que se dan en una contigüidad temporal. En consecuencia, el rol del docente es comunicar un conocimiento concreto de forma clara y estructurada, explicar soluciones correctas, confrontar a los estudiantes con problemas claros y resolubles, así como mantener el orden y la disciplina en el salón de clases. En contraste, la visión constructivista pone su atención en los procesos cognitivos individuales e interpersonales en la construcción del conocimiento, donde el estudiante es un participante activo en el proceso de aprendizaje. De esta manera, la labor del docente es facilitar la propia búsqueda de los estudiantes, permitirles participar en la solución de problemas y posibilitar una coparticipación en las actividades de enseñanza. Por consiguiente, se enfatiza el desarrollo de los procesos de pensamiento y razonamiento más que la adquisición del conocimiento concreto.

TABLA 1. Reactivos que conforman las escalas de creencias pedagógicas: transmisión directa y constructivista

Escala	Reactivos
Instrucción directa	Los profesores eficaces/buenos demuestran la manera correcta de resolver un problema La enseñanza debe basarse en problemas con respuestas claras y precisas y en ideas que los alumnos puedan captar con rapidez Lo que aprenden los alumnos depende de los conocimientos previos que posean, por eso la enseñanza de hechos resulta indispensable Por lo general, el aprendizaje eficaz requiere una clase con un ambiente en silencio
Visión constructivista	Mi papel como profesor es ayudar a los alumnos a realizar sus propias investigaciones La mejor forma de aprender para los alumnos es que ellos encuentren las soluciones a los problemas por sí mismos Debe permitirse a los alumnos que encuentren soluciones a los problemas de carácter práctico por sí mismos antes de que el profesor les enseñe la manera de resolverlos Los procesos relacionados con el pensamiento y el razonamiento son más importantes que el contenido específico del currículo

Fuente: Backhoff, Andrade, Bouzas, Santos y Santibáñez (2009).

Las creencias sobre los enfoques de la instrucción directa y la visión constructivista se evaluaron con dos escalas conformadas por seis preguntas cada una. En ambos casos se utilizó una escala Likert² de cuatro puntos: «totalmente en desacuerdo», «en desacuerdo», «de acuerdo» y «totalmente de acuerdo». En la tabla 1 se muestran los reactivos utilizados en estas dos escalas.

Es importante señalar que en el estudio internacional de TALIS se puso a prueba la comparabilidad cultural —o invariancia— de las escalas utilizadas para medir las creencias, las prácticas y las actitudes de los docentes de los 23 países que participaron en este estudio; por ello, se aplicó el método de análisis de varianza confirmatorio. Los resultados de estos análisis no confirmaron la condición de invariancia de algunas de estas escalas, lo que quiere decir que su comportamiento varía de un país a otro³. Consecuentemente, TALIS recomendó para estas escalas no realizar comparaciones de las puntuaciones promedio de los países. Por esta razón, los análisis internacionales pusieron su atención en el comportamiento general de los países en dichas escalas, más que en la comparación de un país con otro (para mayor información, consultar: OCDE, 2009)⁴.

Tomando esto en consideración, TALIS decidió que los resultados de estas escalas se reportaran en puntuaciones *ipsativas*, las que se utilizan para corregir la tendencia global de aceptar o rechazar cualquiera de las creencias estudiadas en TALIS y para ayudar a corregir el sesgo de las respuestas culturales de un país. Este tipo de puntuaciones se calculan sustrayendo el promedio (o media) de todos los reactivos que conforman una escala de la puntuación individual de cada reactivo de la escala⁵. Es decir, es una medida de variación que mide cuán lejos o cerca se ubican las respuestas de un país (u otra unidad de agregación) respecto a la puntuación media. Las puntuaciones positivas indican que una creencia (o una práctica) recibe un apoyo o soporte fuerte en un país (modalidad educativa

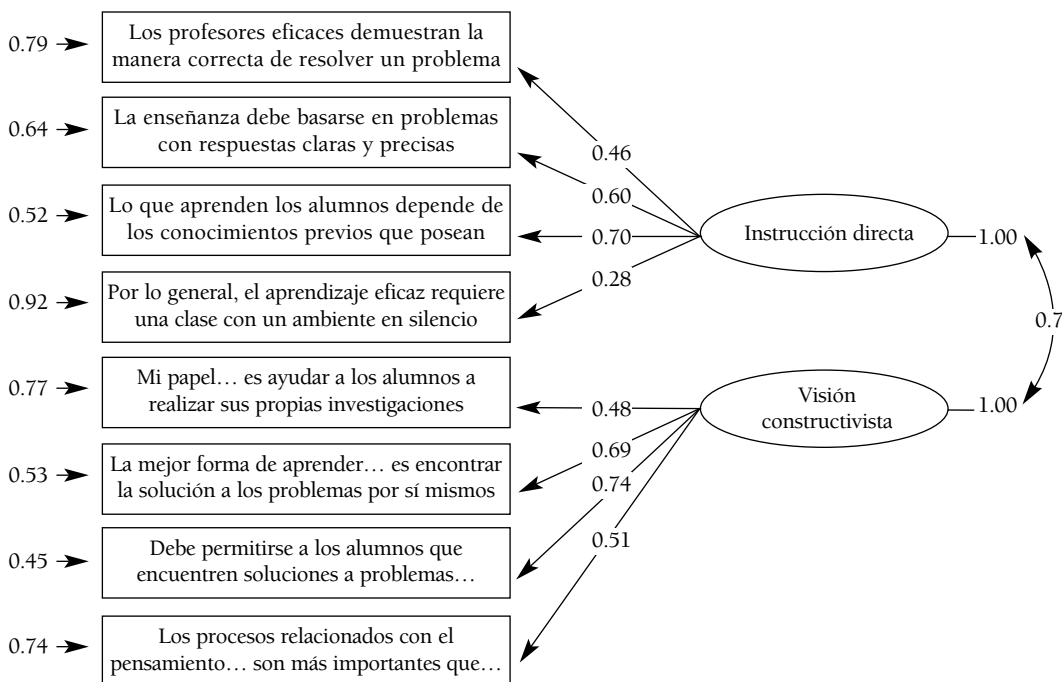
o Estado), mientras que los valores negativos indican un apoyo relativamente débil.

Como ya se mencionó, 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 teórica (o reproducida) que será comparada con la matriz empírica derivada de los datos observados. Para el constructo *creencias docentes de los profesores mexicanos de tercero de secundaria*, la representación gráfica se muestra en la figura 1.

La flecha bidireccional representa covarianza entre las dos variables latentes. Las flechas unidireccionales representan las líneas de influencia de las variables latentes (tipos de creencias pedagógicas) sobre las respectivas variables observadas (reactivos del cuestionario). 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, los coeficientes y flechas unidireccionales que aparecen a la izquierda de los recuadros representan la influencia de fuentes externas de variabilidad única (o idiosincrática) asociada a cada variable.

En la figura 1 podemos observar que la variable latente *instrucción directa* ejerce una influencia diferencial ponderada con coeficientes de 0.46, 0.60, 0.70 y 0.28, respectivamente, sobre las cuatro variables observadas (reactivos), con las cuales se intenta medir este tipo de creencia docente: 1) *Los profesores eficaces/buenos demuestran la manera correcta de resolver un problema*, 2) *La enseñanza debe basarse en problemas con respuestas claras y precisas y en ideas que los alumnos puedan captar con rapidez*, 3) *Lo que aprenden los alumnos depende de los conocimientos previos que posean, por eso la enseñanza de hechos resulta indispensable* y 4) *Por lo general, el aprendizaje eficaz requiere una clase con un ambiente en silencio*.

FIGURA 1. Representación gráfica del modelo de medición original de creencias docentes



Chi-Square = 5445.88, df = 19, P-value = 0.00000, RMSEA = 0.068

Es decir, en la medida en que un profesor está de acuerdo con los cuatro reactivos anteriores, manifiesta proporcionalmente su creencia de que la *instrucción directa* es el mejor método de enseñanza. Cuanto mayor sea el coeficiente, mayor será la asociación entre el contenido de un reactivo y la creencia que lo sustenta. Por ejemplo, el coeficiente más alto en este grupo de reactivos es de 0.70 y corresponde a la afirmación de que el aprendizaje depende del conocimiento previo y que la enseñanza de hechos resulta indispensable. A su vez, el coeficiente más bajo (0.28) indica que una clase con un ambiente en silencio es también importante para el aprendizaje eficaz, pero es menos importante que el reactivo anterior con relación a la creencia sobre la *instrucción directa*.

Para interpretar estos coeficientes hay que tener en cuenta que, siguiendo la teoría SEM, cuando

la creencia en las bondades de la *instrucción directa* aumenta en una unidad, las respuestas a cada uno de los cuatro reactivos que componen esta escala aumentarán proporcionalmente en 0.46, 0.60, 0.70 y 0.28 puntos; lo que revela la magnitud y dirección de la relación que existe entre la variable latente (creencia docente) y cada una de las variables observadas (reactivos).

En cuanto a las relaciones entre la *visión constructivista* y los cuatro reactivos con los que se intenta medir esta variable, el modelo encuentra los siguientes coeficientes: 0.48 para *Mi papel como profesor es ayudar a los alumnos a realizar sus propias investigaciones*; 0.69 para *La mejor forma de aprender para los alumnos es que ellos encuentren las soluciones a los problemas por sí mismos*; 0.74 para *Debe permitirse a los alumnos que encuentren soluciones a los problemas de carácter práctico por sí mismos antes de que el profesor les enseñe la*

manera de resolverlos y; 0.51 para *Los procesos relacionados con el pensamiento y el razonamiento son más importantes que el contenido específico del currículo*. Estos resultados indican que, a medida que la *visión constructivista* del docente aumenta en una unidad, las respuestas a los cuatro reactivos aumentan en la misma proporción que tengan sus coeficientes de influencia.

Por otro lado, un dato importante en este primer modelo es el relacionado con el término de covarianza entre los dos variables latentes o tipos de creencias, que es de 0.7. Esto se interpreta en el sentido de que las creencias docentes *instrucción directa* y *visión constructivista* (medidas con estos ocho reactivos) comparten elementos comunes, de tal forma que no se pueden considerar constructos totalmente distintos y excluyentes. En otras palabras, hay docentes con creencias más apegadas a la *visión constructivista* que también sustentan opiniones favorables a uno o más de los reactivos propios de la *instrucción directa* y viceversa.

Finalmente, en la sección izquierda de la figura se representa la influencia de fuentes externas de variabilidad que se asocian a cada uno de estos reactivos y que no están contempladas en el modelo. Estas influencias pueden ser, entre otras, las distintas interpretaciones de los docentes al contenido de los reactivos, de tal manera que la idiosincrasia de los profesores respecto a las afirmaciones también juega un papel importante en la interpretación de los docentes sobre el contenido de los reactivos y la forma de responderlos. Esto a su vez indica que, si bien los reactivos se relacionan con la dimensión medida, por sí solos no constituyen medidas puras de la variable latente o creencia pedagógica que se intenta medir. Para apreciar estas influencias, debe considerarse que cuanto menor sea el coeficiente de variabilidad externa de los reactivos, mayor precisión tendrán para medir la creencia pedagógica correspondiente.

Ahora bien, la prueba definitiva de la estructura de esta escala consiste en establecer en qué

medida este modelo teórico, propuesto por la OCDE para evaluar los dos tipos de creencias docentes, es congruente con los datos empíricos que proporcionaron los miles de docentes mexicanos que respondieron el cuestionario. Para ello hay que evaluar la *bondad de ajuste* entre la matriz derivada de los datos y la matriz reproducida por el modelo, por lo que la aproximación entre los elementos de ambas matrices se somete a prueba.

Como ya se mencionó, la metodología SEM ha propuesto una serie de índices y criterios de bondad de ajuste para determinar si la condición de aproximación entre las matrices se sostiene con rigor suficiente para concluir que el modelo de medida propuesto y los datos obtenidos de la aplicación muestral de la escala ajustan entre sí.

Los índices de bondad de ajuste existentes son muy diversos, pero todos tienen en común la siguiente característica: son una medida de la diferencia absoluta observada entre cada uno de los elementos de la matriz de covarianzas derivada de los datos empíricos y la matriz reproducida a partir del modelo de medición hipotetizado en la estructura conceptual de las escalas. Cuando las diferencias detectadas entre los elementos son mínimas —dentro de criterios de límite previamente establecidos— se concluye que las estructuras de ambas matrices se aproximan y por lo tanto se concluye que los datos y el modelo teórico coinciden.

El primer índice de bondad de ajuste es el estadígrafo χ^2 (chi cuadrada) que se propone para representar el punto de mejor ajuste o discrepancia mínima entre las matrices comparadas (Blalock, 1964). Como se puede observar en la figura 1, el valor de χ^2 resultó ser igual a 5.445,88 ($p=0.000$). 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, la muestra de docentes evaluada fue de 62.650. Cuando se presenta esta situación, se matiza 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) (Steiger, 1990). En la misma figura 1 se observa que el valor del RMSEA fue de 0.068, lo que significa que el modelo de medición del cuestionario de alumnos y la estructura de covarianzas de las respuestas de los estudiantes tiene un ajuste razonable, aunque no óptimo (que debería ser menor a 0.05 y en todos los casos nunca mayor de 0.10).

Por esta razón, deben incorporarse otras evidencias que confirmen los resultados obtenidos. Para ello se recurrió a interpretar tres índices más: el índice de ajuste comparativo (CFI), el índice de ajuste incremental (IFI) y la raíz cuadrática media de residuales (RMSR)⁶. El valor del CFI del modelo evaluado fue de 0.96, lo que significa que el modelo de medición del cuestionario y la estructura de los datos se aproximan a un buen ajuste, ya que logran el mínimo aceptable de 0.90. Asimismo, el modelo tuvo un IFI de 0.96, que está muy cerca de su parámetro óptimo que es de 1. Finalmente, el valor de RMSR fue de 0.053, que es ligeramente superior al valor requerido (0.05) para considerar que el modelo tiene un ajuste razonable

En síntesis, con los valores de bondad de ajuste observados, se concluye que el modelo conceptual original y los datos observados no presentan un ajuste suficiente, aunque se aproximan a los criterios de un ajuste razonable. Dada esta

situación, es pertinente re-especificar el modelo original y con ello mejorar los índices de bondad de ajuste del modelo teórico, como se muestra continuación.

Se aclara que este ejercicio de re-especificación cancela el carácter confirmatorio de este análisis y lo convierte en un ejercicio de carácter exploratorio para el modelo en cuestión. Esta búsqueda de mejor ajuste (specification search) se basa en sugerencias que parten de evidencia estadística proporcionada por la prueba de multiplicadores de LaGrange. Esta indica reducciones posibles en los valores iniciales de discrepancia en la χ^2 al agregarse relaciones adicionales no consideradas en el modelo original. Se aclara también que esta re-especificación no está basada en teoría sustantiva, sino que en todo caso, se ofrece como indicio de que la teoría sustantiva original, donde las dos sub-dimensiones *instrucción directa* y *visión constructivista* están separadas o claramente delimitadas en las creencias de los docentes es cuestionable.

La figura 2 muestra los resultados de un modelo de medición re-especificado que cumple mejor con los criterios de ajuste que exige la metodología SEM. El modelo contiene elementos nuevos y relaciones adicionales entre las dimensiones y los reactivos, así como entre las fuentes externas de varianza. Debe observarse que la covarianza entre las dos dimensiones es ahora de 0.87, contra 0.70 del modelo original. Esto nos indica de que el modelo re-especificado establece una relación aún más estrecha entre las variables latentes: *instrucción directa* y *visión constructivista*, por lo que se puede decir que comparten más elementos en común y se hacen menos excluyentes.

Por otro lado también se pueden observar dos nuevas relaciones entre las creencias docentes y los reactivos. El modelo establece que la creencia en la *instrucción directa* tiene un vínculo negativo considerablemente grande (-1.70) con un reactivo que pertenece a la *visión constructivista*:

la mejor forma de aprender para los alumnos es que ellos encuentren las soluciones a los problemas por sí mismos. El símbolo negativo implica que a medida que el docente cree más en la *instrucción directa*, menos está de acuerdo en que los alumnos deben encontrar soluciones a los problemas por sí mismos, lo cual tiene mucho sentido.

Asimismo, el nuevo modelo establece que la *visión constructivista* tiene una influencia positiva moderada (0.32) con un reactivo que pertenece a la instrucción directa: *la enseñanza debe basarse en problemas con respuestas claras y precisas y en ideas que los alumnos puedan captar con rapidez*. Además de que esta relación no estaba contemplada en el modelo original, no es de esperarse que los reactivos que miden un tipo de creencia docente tengan una relación positiva con una creencia opuesta. Contrario al caso anterior, esta relación no tiene mucho sentido.

Esta nueva relación, como la anterior, explica en parte el incremento en las covarianzas entre los dos tipos de creencias, lo que indica que los docentes tienden a compartir opiniones de dimensiones que conceptualmente representan puntos de vista contrarios sobre la enseñanza.

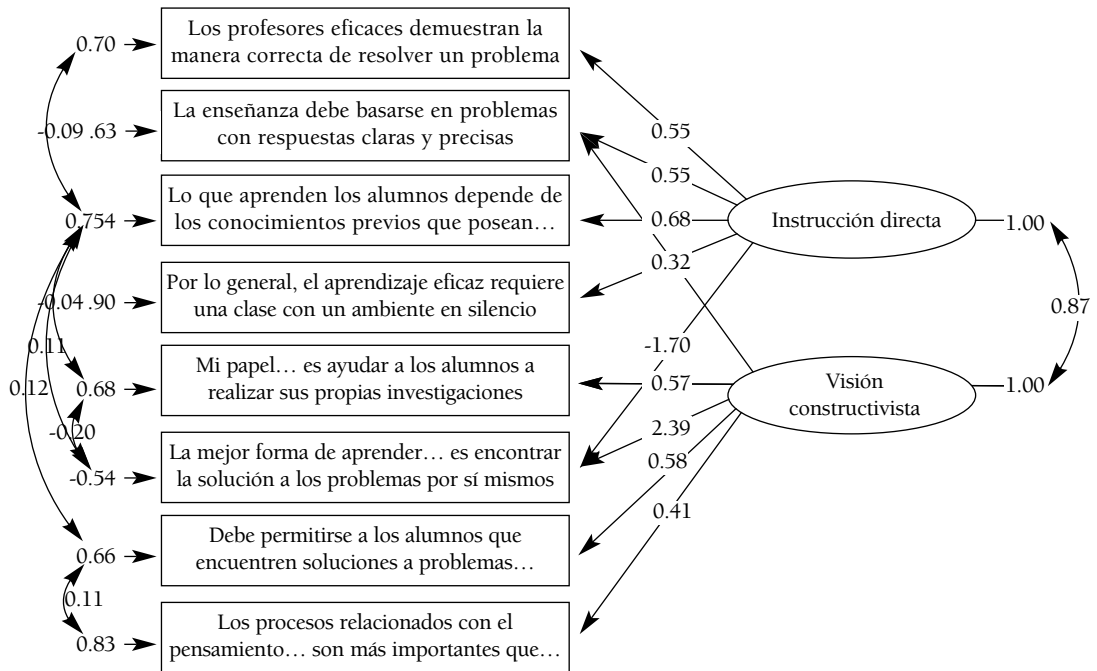
Hay que notar también que ha cambiado la intensidad de las relaciones entre las dimensiones y las variables observadas, en algunos casos modestamente y en otros considerablemente. Un caso importante es el gran incremento en la relación de las creencias constructivistas con el reactivo que hace alusión a *la mejor forma de aprender para los alumnos es que ellos encuentren las soluciones a los problemas por sí mismos*, que antes era de 0.74 y que en el nuevo modelo es de 2.39. Esto ocurre debido a que como los reactivos, por sí mismos, no son medidas puras de las dimensiones, al establecerse nuevas relaciones entre las fuentes externas de varianza, en este caso, se revela con mayor intensidad la relación entre la *visión constructivista* y el

reactivo, cuyo contenido lo representa *La mejor forma de aprender* [...].

En el extremo izquierdo de la figura se observan flechas bidireccionales que indican covarianza entre las fuentes externas de dos o más reactivos. Por ejemplo, el primer reactivo (*los profesores eficaces...*) y el tercero (*lo que aprenden los alumnos depende...*) presentan una covarianza de -0.09 lo que indica que las opiniones frente a estos dos reactivos comparten una fuente de varianza en común que los impacta levemente en sentido opuesto (debido al signo negativo). Lo anterior se explica debido a que los dos reactivos no son medidas puras de la dimensión que miden (*instrucción directa*) y a que comparten una influencia externa común que no estaba contemplada en el modelo original. Este mismo análisis se puede hacer de las demás covarianzas asociadas a las fuentes externas de variabilidad que aparecen en el extremo izquierdo de la siguiente figura.

Por otra parte y como prueba adicional se comprueba la significatividad estadística de los coeficientes de influencia estimados a partir del modelo empírico. Se realizan pruebas Z seleccionando previamente un valor de α , donde suele adoptarse el valor de 1.96 correspondiente a $\alpha = .05$. O bien, como es el caso de las pruebas empleadas en este estudio el valor crítico es de 2.58 correspondiente a $\alpha = .01$. Para realizar las pruebas Z individuales se divide cada estimado sobre su respectivo error estándar. Si el cociente obtenido es mayor que un valor crítico, se concluye que el coeficiente en cuestión es distinto a \emptyset en la población. En el presente estudio todos los coeficientes de los modelos finales resultaron ser estadísticamente significativos. El anexo a contiene los resultados individuales de las pruebas Z realizadas sobre los coeficientes estimados en cada sección de este trabajo. En sentido estricto estas pruebas de significancia son pruebas t, sin embargo debido al tamaño de muestra en este estudio ($n= 62.650$) se convierten a pruebas bajo la distribución Z.

FIGURA 2. Representación gráfica del modelo de medición re-especificado de Creencias docentes



Chi-Square = 2661.66, df = 11, P-value = 0.00000, RMSEA = 0.062

Como resultado de la re-especificación del modelo, se obtiene un índice de modificación que reduce el valor de χ^2 2,661.66. Esto representa una reducción sustancial de la medida de discrepancia del modelo original ($\chi^2 = 5,445.88$ contra $\chi^2 = 2,661.66$ que es igual a una reducción $\Delta\chi^2$ de 2,784.22). Esta reducción del valor de χ^2 mejora sustancialmente el índice global de bondad de ajuste del modelo empírico. En la tabla 2 se muestran los cinco índices analizados, donde se podrá observar que solo en el indicador de χ^2 y en el RMSEA el nuevo modelo no cumplió totalmente con los criterios establecidos, por lo que consideramos que el modelo re-especificado solo cumple parcial o marginalmente con los criterios de bondad de ajuste de la metodología SEM.

Esto último se basa en que si bien se aceptan valores de RMSEA menores a .06 (Hu y Bentler, 1999) como indicativos de buen ajuste, se debe notar que aun bajo la re-especificación exploratoria el valor obtenido de .062 rebasa este criterio. Además, al construirse un intervalo de confianza de 90% alrededor del valor obtenido, se tiene que este fluctúa entre .06 y .064. También es de destacarse que esta relativa mejoría en el ajuste de logra estableciendo múltiples covarianzas entre los términos de error de las variables observadas, lo cual sugiere que están presentes fuentes de variabilidad entre ellas que no se contemplan en el modelo original. Esto refuerza la conclusión de que el ajuste entre este modelo y los datos de la muestra con representatividad nacional es marginal en el mejor de los casos.

TABLA 2. Índices de bondad de ajuste del modelo de medición re-especificado de creencias docentes

Índice	Valor obtenido	Criterio límite	Interpretación
X ²	0.000	P > 0.05	No cumple
RMSEA	0.062	< 0.05	No cumple
CFI	0.98	> 0.90	Cumple
IFI	0.98	> 0.90	Cumple
SMRS	0.04	< 0.05	Cumple

Prácticas pedagógicas

El estudio TALIS se propuso identificar diferentes tipos de enseñanza que se utilizan en el aula, más que identificar la *mejor* práctica docente. Partiendo de esta perspectiva se construyeron tres escalas con diversos reactivos cada una, en las que se les pidió a los docentes

estimar la frecuencia con que realizan diversas actividades dentro del salón de clases. Para ello, se utilizaron los siguientes rangos: «nunca o casi nunca», «aproximadamente en una cuarta parte de las clases», «aproximadamente en la mitad de las clases», «aproximadamente en unas tres cuartas partes de las clases» y «casi en todas las clases».

TABLA 3. Reactivos que conforman las escalas de prácticas pedagógicas: estructuradas, orientadas al estudiante y actividades enriquecidas

Escala	Reactivos
Prácticas estructuradas	Establezco de forma explícita los objetivos de aprendizaje
	Corrijo con los alumnos las tareas que han realizado
	Al comienzo de la lección presento un breve resumen de la lección anterior
	Corrijo los cuadernos de ejercicios de mis alumnos
	Compruebo, a través de preguntas, si el tema se ha entendido o no
Prácticas orientadas en el estudiante	Los alumnos trabajan en pequeños grupos para alcanzar una solución conjunta a un problema o tarea
	Proporciono actividades distintas a los alumnos con problemas de aprendizaje y a aquellos que pueden avanzar con mayor rapidez
	Pido a mis alumnos que propongan o me ayuden a planificar las actividades o temas de clase
	Los alumnos trabajan en grupos organizados según sus capacidades
Actividades enriquecidas	Los alumnos trabajan en proyectos cuya finalización requiere al menos una semana
	Los alumnos elaboran un producto que será utilizado por otras personas
	Pido a mis alumnos que realicen un trabajo por escrito en el que expongan sus ideas y razonamientos de forma relativamente extensa
	Los alumnos mantienen un debate y defienden un determinado punto de vista que puede no ser el propio

Fuente: Backhoff, Andrade, Bouzas, Santos y Santibáñez (2009).

En la tabla 3 se presentan los contenidos de las tres escalas utilizadas para evaluar las prácticas pedagógicas de los docentes: prácticas estructuradas, prácticas orientadas al estudiante y actividades enriquecidas. La primera fue construida con cinco reactivos, mientras que la segunda y la tercera se construyeron con cuatro cada una. En este recuadro se muestran los reactivos utilizados en dichas escalas.

Como se podrá apreciar en esta tabla, la escala de prácticas estructuradas mide el grado en que los docentes diseñan sus clases, exponen la información con claridad y de forma gradual

y los estudiantes conocen las metas de aprendizaje. Por su parte, la escala de prácticas orientadas al estudiante mide el grado en que los docentes permiten que los estudiantes participen activamente en su aprendizaje y trabajen en grupo. Finalmente, la escala de actividades enriquecidas mide el grado en que el docente asigna a sus estudiantes proyectos, ensayos y ciertos productos escolares de mayor alcance y profundidad.

TALIS solicitó a los docentes que reportaran sus prácticas de enseñanza para una asignatura específica, la cual se definió (arbitrariamente)

FIGURA 3. Representación gráfica del modelo de medición original de Prácticas pedagógicas



Chi-Square = 8649.02, df = 62, P-value = 0.00000, RMSEA = 0.047

como la primera clase que el profesor normalmente imparte los días martes después de las 11 a.m. (OCDE, 2011). De esta manera se consiguió que el profesor no eligiera alguna asignatura en particular y, con ello, evitar sesgos en los resultados. Hay que decir que los docentes pudieron responder sobre cualquiera de las asignaturas académicas (por ejemplo: matemáticas, lengua española, ciencias naturales) o sobre otro tipo de asignaturas (artes, tecnologías o deportes).

Los resultados del modelamiento SEM para las prácticas pedagógicas se muestran en la figura 3 y en la tabla 4, donde se puede observar que el valor de χ^2 resultó igual a 8,649.02 ($p=0.000$), valor de p que es menor a 0.05, lo que indica que la diferencia entre las matrices de covarianzas comparadas es estadísticamente significativa. Sin embargo, como ya se mencionó, este resultado no es definitivo, debido a su sensibilidad cuando el tamaño de la muestra es muy grande.

Como se podrá apreciar en la tabla 4 el modelo original de los tres tipos de *prácticas pedagógicas* evaluadas presenta una bondad de ajuste muy aceptable, ya que su RMSEA fue de 0.047, su IFI y CFI fueron cada uno de 0.98 y su SMRS fue de 0.028. Es decir, solo en el indicador de χ^2 del modelo original no cumplió con los criterios establecidos, por lo que consideramos que este modelo cumple óptimamente con los criterios de bondad de ajuste de la metodología SEM y,

en consecuencia, no se requiere re-especificar el modelo. No obstante, como se muestra en la figura 2, hay que destacar que los tres tipos de prácticas de enseñanza covarían, especialmente las prácticas estructuradas y las prácticas orientadas hacia el estudiante, cuyo nivel de covarianza es de 0.76, lo que indica dichas prácticas pedagógicas son concurrentes entre las opiniones de los docentes y por lo tanto estas características en la práctica no son totalmente excluyentes.

Cooperación profesional entre profesores

Los cuestionarios de TALIS cubrieron varios aspectos de la cooperación entre docentes, entendiendo esta actividad como el trabajo que realizan equipos de profesores con el propósito de mejorar los procesos y resultados educativos de las escuelas (Clement y Vandenberghe, 2000). Para ello, TALIS distinguió dos tipos diferentes de cooperación entre docentes, a los que llamó: *intercambio y colaboración docente* y *colaboración para el desarrollo profesional*. El primero implica un tipo de colaboración más tradicional, mientras que el segundo requiere de una colaboración más moderna e innovadora. Para evaluar cada tipo de colaboración se construyeron dos escalas, de cinco preguntas cada una, con los reactivos que se muestran en la tabla 5 y que fueron respondidos por los profesores en una escala Likert de seis puntos: «nunca», «menos de una vez al año», «una vez al año», «3 ó 4 veces al año», «mensualmente» y «semanalmente».

TABLA 4. Índices de bondad de ajuste del modelo de medición original de prácticas pedagógicas

Índice	Valor obtenido	Criterio límite	Interpretación
χ^2	0.000	$P > 0.05$	No cumple
RMSEA	0.047	< 0.05	Cumple
CFI	0.980	> 0.90	Cumple
IFI	0.980	> 0.90	Cumple
SMRS	0.028	< 0.05	Cumple

TABLA 5. Reactivos que conforman las escalas de cooperación entre profesores: intercambio y colaboración para la enseñanza y colaboración para el desarrollo profesional

Escala	Reactivos
Intercambio y colaboración para la enseñanza	Tengo voz y voto en la selección de los medios de enseñanza: libros de texto, cuadernos de ejercicios...
	Intercambio materiales didácticos con los compañeros
	Asisto a reuniones docentes del grado/ asignatura en las que imparto clase
	Me aseguro de que haya normas comunes en las evaluaciones para evaluar el progreso de los alumnos
	Participo en reuniones sobre los progresos relativos al aprendizaje de determinados alumnos
Colaboración para el desarrollo profesional	Imparto clase de forma conjunta, en equipo, en la misma aula
	Participo en actividades de aprendizaje profesional: supervisión de equipos...
	Observo las clases de otros profesores y les realizo observaciones
	Participo en actividades conjuntas con distintos grupos/asignaturas: proyectos...
	Comento y coordino las tareas de las distintas asignaturas

Fuente: Backhoff, Andrade, Bouzas, Santos y Santibañez (2009).

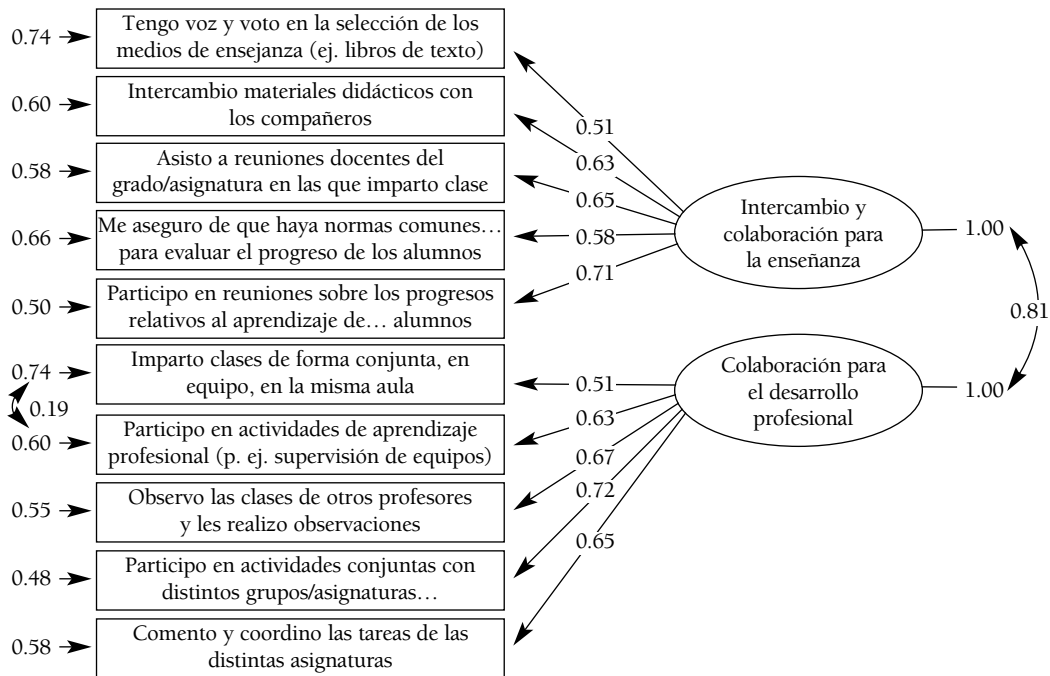
De esta manera, la escala de intercambio y colaboración docente buscó evaluar las siguientes actividades del profesor: elegir medios de enseñanza, intercambiar material didáctico, asistir a reuniones sobre la asignatura que imparte, utilizar normas comunes para evaluar el aprovechamiento y monitorear el progreso educativo de los estudiantes. Por su parte, la escala de colaboración para el desarrollo profesional se conformó con algunas de las siguientes actividades docentes: impartir clases conjuntamente con otros colegas, participar en la supervisión de otros profesores, observar y ser observado en la actividad docente, coordinarse para asignar tareas escolares y participar en proyectos escolares.

Es importante decir que las actividades que conforman cada una de las dos escalas están altamente interrelacionadas. Por ello, es de esperar que, aunque los docentes se involucren más en un tipo de cooperación que en otro lo pueden hacer en ambos. Como en los casos de las creencias docentes y las prácticas pedagógicas, esta escala no implica necesariamente posiciones opuestas y excluyentes.

En la figura 4 y en la tabla 6 se muestran los resultados del modelamiento SEM para el constructo de cooperación profesional entre profesores. En este caso solo se muestran los resultados del modelo re-estructurado, el cual es casi idéntico al modelo original propuesto por la OCDE, con la única variante de que el nuevo modelo agrega una covarianza mínima entre dos reactivos que pertenecen a la dimensión *colaboración para el desarrollo profesional*.

La tabla 6 muestra los índices de bondad de ajuste de este modelo de colaboración docente, donde se puede apreciar que la χ^2 tuvo un valor de 3083.35 ($p=0.000$), el RMSEA fue de 0.038, el IFI y CFI fueron cada uno de 0.99 y el SMRS resultó de 0.028. Es decir, este modelo cumple con cuatro de los cinco índices de bondad de ajuste y, como en los casos anteriores, el indicador de χ^2 del modelo no cumple con el criterio establecido. Hay que puntualizar que las dos dimensiones de colaboración profesional covarían (0.81) y que todos los reactivos tienen asociadas fuentes externas de variabilidad, lo

FIGURA 4. Representación gráfica del modelo de medición re-especificado de colaboración entre profesores



Chi-Square = 3083.35, df = 33, P-value = 0.00000, RMSEA = 0.038

TABLA 6. Indicadores de ajuste del modelo de medición re-especificado de colaboración entre profesores

Índice	Valor obtenido	Criterio límite	Interpretación
χ^2	0.000	> 0.50	No cumple
RMSEA	0.038	< 0.05	Cumple
CFI	0.99	> 0.90	Cumple
IFI	0.99	> 0.90	Cumple
SMRS	0.28	<0.05	Cumple

que indica que dichas formas de colaboración profesional no son totalmente excluyentes y que la forma de medirlas a través de los reactivos analizados tiene ciertos grados de imprecisión.

Discusión

Este trabajo se propuso investigar la validez de constructo de tres escalas docentes utilizadas en el estudio TALIS-2008, cuyo objetivo se centró

en conocer (y comparar) diversas características de las escuelas secundarias de 23 países (OCDE, 2009).

Es común que en los estudios internacionales (al menos los de la OCDE), los instrumentos de medición que se utilizan (pruebas de aprendizaje y cuestionarios) se diseñen y construyan originalmente en inglés (y en francés) y, posteriormente, se traduzcan y se adapten a otros idiomas. Desgraciadamente, por motivos económicos y de tiempo, en muchas ocasiones, este proceso no se hace con el cuidado necesario para asegurar una validez cultural de los resultados obtenidos (Solano-Flores, Backhoff y Contreras-Niño, 2009).

Este parece ser el caso del estudio TALIS-2008, en que se reportó que las tres escalas docentes, objetos de este estudio, no mostraron la invarianza necesaria para poder realizar las comparaciones entre los países participantes, razón por la cual se tuvieron que utilizar puntuaciones ipsativas para poder analizar las respuestas de los profesores al interior de los países (OCDE, 2009).

Esta situación motivó estudiar la validez de constructo de las tres escalas docentes para el caso de los profesores mexicanos, cuya muestra no solamente fue representativa del país sino también de sus entidades federativas. Con el uso de la modelización de ecuaciones estructurales

se contrastó la estructura conceptual de cada escala con su estructura empírica (obtenida con las respuestas de los profesores).

Los resultados obtenidos muestran que, para el caso de México, la escala de creencias docentes aun después de haber sido reestructurada, no tiene los niveles de ajuste necesarios para considerarla válida. Por su parte las escalas de prácticas pedagógicas y colaboración profesional mostraron tener los niveles de ajuste necesarios para considerar válidas sus estructuras conceptuales.

Aunque se desconoce la razón de los pobres niveles de ajuste de la escala de creencias docentes, es pertinente considerar que se puedan deber a tres grandes factores: problemas conceptuales de la teoría sustantiva (selección de reactivos), problemas de origen (mal diseño de los reactivos) y problemas de adaptación y traducción de los ítems (reactivos culturalmente deficientes). Los tres factores apuntan a un problema de validez, inherentes a los estudios internacionales. Por tal razón, independientemente del factor responsable, es recomendable poner la mayor atención posible a la validez de constructo de los instrumentos que se utilizan en un estudio que compara los resultados de diversas naciones antes de implementarlo a gran escala: los recursos y esfuerzos que se destinen a ello podrán asegurar la validez de las comparaciones internacionales, que es el propósito central de este tipo de estudios.

ANEXO A

Significatividad de los coeficientes en el modelo de medición creencias docentes*

	INSTD	VCONS
29A	0.55 (0.01)	--
	86.90	
29G	0.55 (0.01)	0.07 (0.01)
	47.53	6.70
29H	0.68 (0.01)	--
	100.51	
29K	0.32 (0.01)	--
	50.81	
29D	--	0.57 (0.01)
		97.06
29F	-1.70 (0.16)	2.39 (0.16)
	-10.74	15.04
29I	--	0.58 (0.01)
		89.88
29L	--	0.41 (0.01)
		69.34

* Se destacan en negritas los valores t / Z superiores al valor crítico de A = .001 que es 2.58.

ANEXO B

Significatividad de los coeficientes del modelo de medición prácticas pedagógicas**

	PRAES	PRAOE	ACTEN
42B	0.67 (0.00)	--	--
	190.28		
42C	0.74 (0.00)	--	--
	237.40		
42H	0.72 (0.00)	--	--
	234.54		
42I	0.74 (0.00)	--	--
	250.62		
42M	0.74 (0.00)	--	--
	226.84		
42D	--	0.71 (0.00)	--
		212.14	
42E	--	0.71 (0.00)	--
		218.81	
42F	--	0.61 (0.00)	--
		157.95	
42N	--	0.64 (0.00)	--
		176.74	
42J	--	--	0.65 (0.00)
			163.33
42O	--	--	0.66 (0.00)
			158.74
42Q	--	--	0.68 (0.00)
			173.78
42S	--	--	0.66 (0.00)
			165.55

** En este caso el error estándar de los estimados es menor que .0001 por lo que los valores de significancia de las pruebas t / Z son mucho mayores que el valor crítico de 2.58.

ANEXO C
Significatividad de los coeficientes del modelo
de medición colaboración docente ***

	INTERC	COLDEP		INTERC	COLDEP
30C	0.51	--			
	(0.00)			30I	0.63
	115.23				(0.00)
30D	0.63	--			167.24
	(0.00)		30J	--	0.67
	167.65				(0.00)
30E	0.65	--		30K	--
	(0.00)				0.72
	162.52				(0.00)
30F	0.58	--		30L	--
	(0.00)				0.65
	148.30				(0.00)
30G	0.71	--			153.05
	(0.00)				
	210.44				
30H	--	0.51			
		(0.00)			
		108.97			

***En este caso el error estándar de los estimados es también menor a .0001 por lo que los valores de significancia de las pruebas t / Z son mucho mayores al valor crítico de 2.58.

Notas

¹ Disponible en el Instituto Nacional para la Evaluación de la Educación: <http://www.inee.edu.mx>

² Escala ordinal que se utiliza para evaluar el grado o nivel en que un individuo está de acuerdo con una afirmación en particular. Por lo general se utilizan cuatro o más categorías que se ordenan en forma ascendente o descendente.

³ Esto puede ocurrir por diferencias culturales, entre otras razones.

⁴ No obstante, al interior de un país es válido realizar comparaciones de los resultados promedio de estas escalas entre distintos grupos de docentes, como es el caso de las modalidades educativas y de las entidades federativas de México.

⁵ Por ejemplo, si el promedio de los reactivos de una escala es igual a «y», la puntuación de cada ítem «x» será igual a «x - y».

⁶ En los tres casos las siglas corresponden al nombre de estos índices en el idioma inglés.

Referencias bibliográficas

- BACKHOFF, E., ANDRADE, E., BOUZAS, A., SANTOS, A. y SANTIBAÑEZ, L. (2009). *Estudio Internacional sobre la Enseñanza y el Aprendizaje (TALIS): Resultados de México*. México, D. F.: Secretaría de Educación Pública.
- BLALOCK, H. M. (1964). *Causal Inference in non-experimental research*. Chapel Hill: University of North Carolina Press.
- CLEMENT, M. y VANDENBERGHE, R. (2000). Teachers' professional development: a solitary of collegial (ad)venture. *Teaching and Teacher Education*, 16, 81-101.
- JORESKOG, K. y SORBOM, D. (1996-2001). *LISREL 8: User's Reference Guide*. Scientific Software International, Lincolnwood, 3.
- HU, L. y BENTLER, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equations Modeling: a Multidisciplinary Journal*, 6 (1), 1-55.
- KANE, M. T. (2001). Current concerns in Validity Theory. *Journal of Educational Measurement*, 38 (4), 319-342.
- MESSICK, S. (1995). Validity of psychological assessment. Validation inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50 (9), 741-749.
- ORGANIZACIÓN PARA LA COOPERACIÓN Y DESARROLLO ECONÓMICO (OCDE) (2009). *Creating effective teaching and learning environments: First results of TALIS*. París: OCDE.
- PETERSEN, P. L., FENNEMA, E., CARPENTER, T. P. y LOEF, M. (1989). Teachers' pedagogical content beliefs in mathematics. *Cognition and instruction*, 6 (1), 1-40.
- RIZOPOULOS, D. (2006). LTM: an R package for Latent Trait Modeling and Item Response Theory Analyses. *Journal of Statistical Software*, 17 (5), 1-25.
- SOLANO-FLORES, G., BACKHOFF, E. y CONTRERAS-NIÑO, L. A. (2009). Theory of Test Translation Errors. *International Journal of Testing*, 9, 78-91.
- STEIGER, J. H. (1990). Structural Model Evaluation and Modification: an interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.

Fuentes electrónicas

- GONZÁLEZ-MONTESINOS, M. J. y BACKHOFF, E. (2010). Validación de un cuestionario de contexto para evaluar sistemas educativos con Modelos de Ecuaciones Estructurales. *Relieve*, 16, 2. <http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_1.htm> [Fecha de consulta: 7/septiembre/2011].

Abstract

Validity evidence of the teacher's questionnaire in the Teaching and Learning International Study (TALIS-2009)

TALIS-2008 is the first study of the OCED with the purpose of knowing and comparing the conditions in which teaching and learning take place in secondary schools (ninth grade), in 23 countries. Teachers' and principals' surveys were used, designed and constructed in English and French, to be translated later into different languages. This study provides evidence of construct validity of three scales: teacher beliefs, teaching practices and teachers' professional cooperation. The teacher's TALIS questionnaires are analyzed through structural equations modeling techniques under confirmatory factor analysis; 62,650 Mexican teachers participated in the study. In the case of Mexico, results show that two teaching scales have acceptable indexes, but this is not the case for the scale of teacher beliefs, whose adjustment levels did not reach minimum desirable levels (even

after being restructured). We conclude that it is important to provide evidence of construct validity of scales, before large-scale studies take place. Finally, the explanatory power of the scales will be essential to provide recommendations for improvement and, consequently, improve the usefulness of this type of assessment.

Key words: TALIS, Survey validity, Secondary schools, Teachers' beliefs, Pedagogical practices, Structural equations modeling, Latent trait modeling, Confirmatory factor analysis, Construct validity.

Résumé

Evidences de validité du questionnaire pour enseignants de l'Enquête Internationale sur l'Enseignement et l'Apprentissage (TALIS-2009)

TALIS-2008 est la première étude de l'OCDE, dont le but est de déterminer et de comparer les conditions scolaires dans lesquelles l'enseignement et l'apprentissage se développent au troisième cours de l'école secondaire dans 23 pays. Des enquêtes ont été faites à des échantillons nationaux d'enseignants et de directeurs d'école, lesquelles ont été originalement conçues en anglais et en français, pour plus tard être traduites aux différentes langues. Cette étude fournit des évidences de validité du construit en trois échelles différentes: les croyances des enseignants, les pratiques pédagogiques et la coopération professionnelle des enseignants. Nous avons utilisé des techniques de Modélisation des Équations Structurelles, dans sa modalité d'Analyse Factorielle Confirmatoire, avec lesquelles nous avons analysé les opinions de 62.650 enseignants mexicains qui ont participé à l'enquête élargie TALIS. Pour le cas du Mexique, les résultats montrent que deux des échelles considérées ont des niveaux admissibles d'ajustement, mais que pour l'échelle des croyances des enseignants, quelques niveaux d'ajustements n'obtiennent pas les résultats désirables (même après sa restructuration). Nous concluons qu'il est important de fournir des évidences de validité du construit dans les différents questionnaires utilisées, avant de faire des études à grande échelle. Pour en finir, le pouvoir explicatif des échelles sera la clé pour pouvoir offrir des recommandations pour son perfectionnement et, en conséquence, pour améliorer l'utilité de ce type d'évaluations.

Mots clés: TALIS, Validité de questionnaires, Secondaire, Croyances des enseignants, Pratiques pédagogiques, Modélisation des équations structurelles, Modélisation de caractéristiques latentes, Analyse factorielle confirmatoire, Validité du construit, Mexique.

Perfil profesional de los autores

Eduardo Backhoff Escudero

Doctor en Educación por la Universidad Autónoma de Aguascalientes. Investigador (desde 1988) y ex director (1993-2003) del Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California. Titular de la Dirección de Pruebas y Medición del Instituto Nacional para la Evaluación de la Educación de México (2004-2008). Ex editor científico de la Revista Electrónica de Investigación Educativa (2009-2011). Asesor de los proyectos PISA-2012 y TALIS-2013 de la OCDE.

Correo electrónico de contacto: backhoff@uabc.mx

Manuel Jorge González-Montesinos Martínez

Doctor en Psicología Educativa por la Universidad de Arizona - Tucson. Investigador del departamento de Ciencias Sociales desde 1991 en la Universidad de Sonora (México). Investigador psicométrica residente en el INEE en 2008. Miembro del Consejo Técnico INEE desde 2009. Investigador invitado en el departamento de Métodos de Investigación y Diagnóstico Educativo (MIDE), facultad de Filosofía y Ciencias de la Educación, Universidad de Valencia, 2009 y 2012.
Correo electrónico de contacto: mgm@caborca.uson.mx