

Escala de Autoeficacia Docente: análisis estructural e invarianza de medición en docentes peruanos de escuelas públicas

Diciembre 2019,
Vol. 11, N°3, 61-72

revistas.unc.edu.ar/index.php/racc

Dominguez-Lara, Sergio^{* a}; Fernández-Arata, Manuela^a; Merino-Soto César^a; Navarro-Loli, Jhonatan S.^a y Calderón-De la Cruz, Gustavo^a

Artículo Original

Resumen

El objetivo fue realizar un análisis estructural y de invarianza de medición respecto al sexo de la *Escala de Autoeficacia Docente* (EAD) en docentes peruanos de escuelas públicas ($n = 347$; 64.3% mujeres; $M_{edad} = 46.96$ años; $DE_{edad} = 9.05$ años). La estructura interna fue analizada con análisis factorial confirmatorio y *exploratory structural equation modeling* (ESEM). El análisis de invarianza de medición fue basado en análisis factorial multigrupo del modelo bifactor, y se estimó la confiabilidad del constructo y de sus puntuaciones. Los resultados muestran que el modelo de tres factores no recibe soporte, mientras que el modelo bifactor es invariante, evidenciando que el constructo es unidimensional en varones (CFI = .983; RMSEA = .051; WRMR = .577) y mujeres (CFI = .986; RMSEA = .041; WRMR = .620). Además, los indicadores de confiabilidad fueron elevados. Se concluye que la EAD cuenta con propiedades psicométricas adecuadas.

Palabras clave: ESEM, autoeficacia, docente, validez, invarianza de medición, educación.

Abstract

Teacher's Self-Efficacy Scale: structural analysis and measurement invariance in Peruvian teachers of public schools. The aim was to analyze the internal structure and measurement invariance of the Teacher's Self-Efficacy Scale (TSES) in Peruvian teachers from public schools ($n = 347$; 64.3% females; $M_{age} = 46.96$ years old; $SD_{age} = 9.05$ years old). The internal structure was analyzed with confirmatory factor analysis and exploratory structural equation modeling (ESEM). The measurement invariance analysis was based on *multi-group factor analysis*, and self-efficacy scores and reliability were estimated. Results show no evidence for the three-factor model, whereas the bifactor model is invariant, suggesting that the construct is one-dimensional in men (CFI = .983; RMSEA = .051; WRMR = .577) and women (CFI = .986; RMSEA = .041; WRMR = .620); also, the reliability indices were strong. We conclude that the TSES has appropriate psychometric properties.

Keywords: ESEM, self-efficacy, teachers, validity, measurement invariance, education.

Tabla de Contenido

Introducción	61
Método	65
Participantes	65
Instrumento	65
Procedimiento	65
Análisis de datos	65
Resultados	66
Discusión	69
Referencias	70

Recibido el 11 de junio de 2019; Aceptado el 23 de septiembre de 2019

Editores: Josefina Rubiales, Paula Abate, Noelia Villareal, Rocío Foltz e Igancio Acuña

Introducción

La autoeficacia en el proceso enseñanza aprendizaje sigue siendo destacada en las investigaciones y aplicaciones prácticas que se realizan con maestros y estudiantes en el contexto educacional por la importancia que tiene en la motivación y el desempeño del estudiante. La autoeficacia puede definirse como las creencias que tienen las personas con respecto a su capacidad para alcanzar las metas y los objetivos

que se proponen (Bandura, 1986). En el ámbito docente se conceptualiza como el nivel de confianza que tiene el maestro en su capacidad para promover el aprendizaje de los estudiantes.

La autoeficacia del docente está referida a las creencias que tiene el maestro con respecto a su propia capacidad para que los estudiantes logren sus metas de aprendizaje y obtengan buenos resultados, incluso en aquellos que son

^a Universidad de San Martín de Porres, Lima, Perú

*Enviar correspondencia a: Dominguez-Lara, Sergio. E-mail: sdominguezmpcs@gmail.com, sdominguezl@usmp.pe

Citar este artículo como: Dominguez-Lara, S., Fernández-Arata, M., Merino-Soto, C., Navarro-Loli, J. & Calderón-De la Cruz, G. (2019). Escala de Autoeficacia Docente: análisis estructural e invarianza de medición en docentes peruanos de escuelas públicas. *Revista Argentina de Ciencias del Comportamiento*, 11(3), 61-72

considerados difíciles o que se encuentran desmotivados (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). Un alto sentido de eficacia en la enseñanza está asociado a comportamientos positivos de los maestros, quienes se muestran más abiertos a nuevas ideas y más comprometidos con su enseñanza, incentivando de esta manera a la mejora en el desempeño de sus estudiantes (Tschannen-Moran & Woolfolk Hoy, 2001).

La escala de autoeficacia docente (EAD), creada y validada a través del análisis factorial por Tschannen-Moran y Woolfolk Hoy (2001), está conformada por tres dimensiones: eficacia percibida en las prácticas instruccionales, eficacia percibida en el manejo del salón de clase y eficacia percibida en el ajuste del estudiante. Estos autores realizaron tres estudios sucesivos con estudiantes de pregrado en entrenamiento docente (*preservice*) y en servicio docente (*inservice*) para brindar las primeras evidencias acerca de las propiedades psicométricas de la EAD. En el primero, se examinó la validez de contenido mediante el análisis lógico del primer banco de 52 ítems, de los cuales se seleccionaron 32. En el segundo estudio, la cantidad de ítems se redujo de 32 a 18, con los cuales se realizó el primer análisis factorial (factorización con ejes principales y rotación varimax). En primera instancia se decidió por una solución de tres factores ortogonales en base al *Scree-test* (*Eficacia percibida en el ajuste del estudiante*, $\alpha = .82$; *Eficacia percibida en las prácticas instruccionales*, $\alpha = .81$; y *Eficacia percibida en el manejo del salón de clase*, $\alpha = .72$), y posteriormente al realizar un análisis factorial de segundo orden siguiendo el método de factorización anterior, destacaron la fortaleza de un factor general, el cual presentó coeficientes de configuración de magnitud moderada a alta (.40 a .70) y un α elevado (.95).

Finalmente, desarrollaron un tercer estudio a fin de mantener la independencia conceptual de los tres factores encontrados inicialmente, incrementando la cantidad de ítems llegando a una versión de 36 ítems, los cuales fueron aplicados a un grupo de 410 personas de las mismas características que los participantes del estudio anterior. Se empleó la factorización por ejes principales con rotación varimax, extrayendo finalmente tres factores, aunque con un número desigual de ítems, por lo que decidieron considerar, para la versión final, los ocho ítems con los coeficientes de configuración de mayor magnitud

en cada factor, llegando a la versión conocida de 24 ítems. De forma similar al segundo estudio, también se apreció la presencia significativa de un factor de segundo orden. En cuanto a la evaluación de la confiabilidad de los puntajes, los coeficientes α fueron de .87 para *Eficacia percibida en el ajuste del estudiante*; de .91 para *Eficacia percibida en las prácticas instruccionales*; y de .90 en *Eficacia percibida en el manejo del salón de clase*, y de .94 para la puntuación total.

Por su parte, Fives y Buehl (2010) contaron con la participación de 102 docentes en actividad y 270 estudiantes universitarios en entrenamiento docente (77.5% mujeres; no hay reporte de la edad de los participantes). Se efectuó un análisis factorial de ejes principales con rotación varimax en cada una de las muestras evaluadas, extrayendo tres factores en base al análisis paralelo. Al observar la matriz de configuración de ambas soluciones de tres factores, se puede apreciar que la estructura factorial de los docentes en actividad se muestra menos compleja, aunque con algunas diferencias con respecto a la estructura original del test, en comparación con la de los estudiantes en entrenamiento, en donde los ítems mostraron cargas factoriales de magnitudes significativa en diversos factores, pero pese a ello no se realizó un reporte formal de complejidad factorial. En cuanto a la confiabilidad de las puntuaciones, los coeficientes α fueron mayores a .80 para las tres dimensiones (*Eficacia percibida en el ajuste del estudiante*, $\alpha = .81$; *Eficacia percibida en las prácticas instruccionales*, $\alpha = .89$; y *Eficacia percibida en el manejo del salón de clase*, $\alpha = .89$),

Duffin, French y Patrick (2012) realizaron un estudio instrumental contando con la participación de 452 estudiantes universitarios en entrenamiento docente (*pre-service*), divididos en dos muestras de 272 (73.5% mujeres; $M_{\text{edad}} = 20.26$) y 180 (57.8% mujeres; $M_{\text{edad}} = 21.70$). Llevaron a cabo un análisis factorial confirmatorio a través del método de estimación de Máxima Verosimilitud con matrices de covarianzas, evaluando dos modelos: unidimensional y de tres factores oblicuos. La evaluación de los modelos evidencia que ambos presentan ajuste adecuado, aunque pudieron observarse correlaciones interfactoriales cercanas a la unidad en ambas muestras, concluyendo finalmente que el modelo de un factor es el que presenta mayor respaldo empírico en la muestra estudiada. Por último, concluyen que en los estudiantes que inician el entrenamiento docente

no se distinguen las tres variables latentes debido a la escasa práctica en el campo, lo que explicaría la elevada complejidad factorial de sus resultados, pero que a medida que el tiempo pasa, aparece tal diferenciación.

Khairani y Razak (2012) desarrollaron un estudio con 191 maestros en entrenamiento y con 122 profesionales de Malasia, encontrado bajo el modelo de Rasch que la escala proporciona estimaciones válidas y confiables de la autoeficacia percibida en la enseñanza del maestro, aunque el comportamiento de 13 ítems fue significativamente diferente en todos los grupos en el análisis de la función del ítem diferencial (DIF), por lo que sugieren un análisis más profundo de estas discrepancias. Igualmente, utilizando el Modelo Rasch (Chang & Engelhard, 2016) reportan, en una muestra de 554 docentes norteamericanos, que la EAD muestra un rango de dificultad de ítems dentro de la dimensión *manejo del aula*, aunque los ítems se muestran invariables con respecto a los años de experiencia docente y el contexto escolar.

En el contexto peruano, existen estudios empíricos que reportan el uso de la versión traducida (Fernández, 2008), mientras que posteriormente Fernández y Merino (2012) realizaron un análisis preliminar de confiabilidad de las puntuaciones de la EAD en 665 docentes de Lima, 313 de primaria (71.6% mujeres) y 352 de secundaria (51.1% mujeres), cuyas edades oscilaron entre 21 y 60 años. Fue analizada su consistencia interna, siendo los coeficientes α de Cronbach, y sus respectivos intervalos de confianza, de magnitud elevada ($\alpha = .90$), pese a que no fue evaluada la dimensionalidad del instrumento mediante procedimientos analítico-factoriales.

Particularmente en el Perú, el estudio de los recursos personales (como la autoeficacia) es aún emergente en el contexto del trabajo (Calderón-De la Cruz, Dominguez-Lara, & Arroyo-Rodríguez, 2018) siendo una de sus principales limitantes, la escasez de instrumentos de medición orientados a su estudio y el EAD puede ser una alternativa adecuada para examinar el grado de eficacia docente; no obstante, es necesario corroborar los procedimientos de validez y confiabilidad, sobre todo porque estas características psicométricas que se utilizan en diferentes países o culturas, son siempre materia de preocupación por los académicos debido a que algunos estudios revelan que su adaptación en contextos particulares arroja

con frecuencia resultados distintos (Khairani & Razak, 2012).

Asimismo, debe señalarse que las investigaciones precedentes sobre la validez de la EAD únicamente consideraron evaluar su dimensionalidad, pero no otras características estructurales que son también relevantes para comprender la relación de los ítems con sus constructos. Por ejemplo, la similitud de las cargas factoriales dentro del mismo factor, la relación de los ítems con las otras dimensiones de la EAD, y su funcionamiento diferencial según sexo y, la ausencia de esta información, podría comprometer su sustancialidad psicométrica (validez y confiabilidad).

En primer lugar, la similitud factorial entre los ítems es una propiedad relevante pues influye en la confiabilidad y en la validez de constructo, específicamente en por lo menos tres aspectos: primero, en la apropiada elección de un modelo de consistencia interna y, concomitantemente, su coeficiente es influenciado por la similitud de los residuales y las cargas factoriales, características que se identifican como *paralelo*, *tau-equivalente* o *congenérico* (Graham, 2006). El modelo que usualmente se prueba en la mayoría de los modelos derivados del análisis factorial (e.g., CFA) es el congenérico, aunque no sea declarado explícitamente (Graham, 2006). Comparado con éste, el modelo paralelo es altamente restrictivo pues requiere la igualdad de cargas factoriales y residuales; mientras que el modelo tau-equivalente restringe la igualdad de las cargas factoriales únicamente. La elección de un coeficiente de confiabilidad apropiado debe considerar estas características estructurales para evitar obtener estimaciones sobre o subestimadas.

En segundo lugar, la relación de los ítems (que pertenecen a una dimensión) con las otras dimensiones no ha sido explorada en los estudios previos, debido a que en los análisis CFA-SEM se impuso la usual restricción a los ítems de estimar libremente la influencia de los factores sobre solo uno de ellos, y restringir a cero la relación con las demás dimensiones. En una situación así, se debe verificar si los ítems también mantienen relaciones sustanciales con otros constructos mediante la evaluación de las cargas cruzadas mediante los métodos apropiados, ya que es posible que el uso de métodos ortogonales en estudios previos estaría enmascarando las correlaciones interfactoriales existentes (e.g., Duffin et al., 2012) y podrían

sugerir la existencia de un factor general que podría ser analizado a través de modelos jerárquicos, cuya interpretación sería más parsimoniosa. De este modo, es relevante implementar métodos que puedan evaluar esa complejidad al interior de la estructura factorial, tales como el *modelamiento exploratorio de ecuaciones estructurales* (Asparouhov & Muthén, 2009), cuyas aplicaciones en Latinoamérica aún son escasas.

En tercer lugar, la ausencia de sesgo ha sido examinada parcialmente mediante métodos de invarianza de medición y dentro del marco CFA-SEM con relación al nivel de enseñanza, aunque no desde un punto de vista de diferencias de sexo. Esto es relevante dado que varones y mujeres difieren en el modo en que experimentan las cuatro fuentes de autoeficacia (Blanco, Ornelas, Aguirre, & Guedea, 2012). En el ámbito docente cabe resaltar la experiencia de dominio (los docentes mejor evaluados, por estudiantes y compañeros, podrían experimentar mayor eficacia), la experiencia vicaria (efectos producidos por las acciones de otros integrantes de la comunidad educativa), la persuasión social (comunicaciones recibidas por otros, como muestras de respeto o felicitaciones), y estados fisiológicos (las personas en general estiman su confianza, en parte, por el estado emocional que experimentan cuando realizan una acción, y en los docentes esas acciones estarían ligadas a cada dimensión que contempla la *autoeficacia docente*). Entonces, si la evaluación realizada en varones y mujeres no presenta evidencia suficiente de invarianza, personas con el mismo nivel *verdadero* de autoeficacia recibirían puntuaciones distintas debido a sesgos asociados a la medición.

La evaluación de la autoeficacia sigue siendo un factor importante en el mundo académico porque es considerado una construcción universal, unidimensional, fiable y homogéneo (Scholz, Gutiérrez, & Schwarzer, 2002), además de tener una influencia fundamental en diversos ámbitos de aplicación por su asociación con la motivación académica y resultados educativos (Deci & Ryan, 2000), lo que brindaría al docente herramientas personales para afrontar los cambios actuales vinculados a contenidos curriculares o sistemas de evaluación, considerando el carácter predictivo de la autoeficacia sobre el éxito en las actividades laborales (Alessandri, Borgogni, Schaufeli, Caprara, & Consiglio, 2015; Mielniczuk & Laguna, 2018).

Entonces, dado que la autoeficacia docente tiene una influencia positiva en la motivación y logros académicos de los estudiantes (Mojavezi & Poodineh, 2012; Tschannen-Moran & Woolfolk Hoy, 2001), lo cual se relaciona significativamente con resultados educativos a través de la persistencia, entusiasmo, compromiso y comportamiento instructivo del maestro (Tschannen-Moran & Woolfolk Hoy, 2001), es de suma importancia disponer de un instrumento de medida con evidencias de validez y confiabilidad, con ítems que puedan indicar de manera explícita y concreta, el nivel de autoeficacia de los docentes a los directivos de las escuelas, por lo que el objetivo del presente estudio fue realizar un análisis estructural de la EAD (Tschannen-Moran & Woolfolk Hoy, 2001k) en docentes de instituciones públicas de Lima.

Es importante llevar a cabo el estudio en este grupo debido a que existe una creciente preocupación por la deficiente formación inicial y en servicio de los docentes de escuelas públicas, a quienes se le atribuye la responsabilidad de los bajos resultados de los estudiantes en las evaluaciones nacionales e internacionales, generando así desconfianza hacia los docentes con la respectiva desmotivación y baja autoestima (Ministerio de Educación, 2010; UNESCO, 2017). Igualmente, a pesar de la mejora de las remuneraciones de los maestros peruanos en los últimos años, éstas se encuentran entre las más bajas de Latinoamérica, teniendo el docente que realizar trabajos paralelos que afectan negativamente su desempeño profesional (Díaz, 2015). Por ello, coexisten los avances y retrocesos en la valoración social y reconocimiento del docente que se traduce no sólo en los bajos salarios sino también en la carencia de incentivos (Rivero, 2007). De este modo, se espera que al contar con esta herramienta será posible brindar a los docentes, la oportunidad de conocer y, de ese modo, procurar la mejora de su nivel de autoeficacia mediante programas de capacitación, ya que podría indicar cuánto de entrenamiento formal y no formal requiere un docente en entrenamiento o profesional. Inclusive, esta medición podría formar parte de la evaluación de competencias, en el proceso de selección, contratación y evaluación de los docentes, aunque para ello haría falta mayor evidencia sobre la relación con el desempeño docente y compromiso con el trabajo.

Método

Participantes

La muestra estuvo constituida por 347 docentes de escuelas públicas de la ciudad de Lima, en su mayor parte mujeres (64.3%), con edades entre 24 y 69 años ($M = 46.96$, $DE = 9.05$). Un porcentaje elevado se encontraba casado al momento de realizar la encuesta (58.8%). No se encontraron diferencias significativas en la edad según el sexo ($t_{(338)} = .200$, $d = .02$).

En cuanto a su formación profesional, la mayoría provenía de universidades nacionales (45.2%), estudió la especialización en educación secundaria (56.8%), y no contaba con estudios de posgrado (60.2%), aunque un porcentaje elevado indicó que le gustaría volver a estudiar (81%) y que recibe capacitación por parte del Ministerio de Educación (84.1%).

Respecto a su ámbito laboral, la mayor parte tenía estabilidad laboral (personal nombrado; 83%), se encontraba bajo el sistema remunerativo denominado *Antigua Ley* (62.8%), y sus años de experiencia oscilaron entre 2 y 45 años ($M = 19.65$, $DE = 7.789$), no encontrándose diferencias con respecto al sexo ($t_{(340)} = .195$, $d = .02$) ni con el nivel de enseñanza ($t_{(336)} = 1.577$, $d = .17$).

Instrumento

Escala de Autoeficacia Docente (EAD; Tschannen-Moran & Woolfolk Hoy, 2001): consta de 24 ítems escalados en formato Likert de cinco puntos (*Desde Nada hasta En gran medida*), distribuidos en tres factores: *eficacia percibida en el ajuste del estudiante*, que evalúa la valoración del docente en su competencia para influir positivamente en el estudiante (8 ítems, e.g. ¿Cuánto puedes hacer para motivar a los estudiantes que muestran poco interés en el trabajo de la escuela?); *eficacia percibida en las prácticas instruccionales*, que evalúa la autovaloración de su capacidad para generar alternativas que produzcan un mejor rendimiento en el estudiante (8 ítems, e.g. ¿Cuánto puedes hacer para adiestrar a los estudiantes a formular buenas preguntas?); *eficacia percibida en el manejo del salón de clase*, que se refiere a la autovaloración con relación a su capacidad para mantener el orden de la clase y conservar la disciplina (8 ítems, e.g. ¿Cuánto puedes hacer para controlar la conducta perturbadora en clase?).

Procedimiento

La presente investigación es de naturaleza instrumental pues se evaluarán parámetros estadísticos para justificar la calidad métrica del instrumento (Ato, López, & Benavente, 2013).

El proceso investigativo se condujo bajo los principios de la declaración de Helsinki (Asociación Médica Mundial, 1964) y del código de ética del Colegio de Psicólogos del Perú (2017). Los docentes que participaron en el estudio fueron informados sobre los fines del mismo y al aceptar completar la escala firmaron un consentimiento informado. Asimismo, se garantizó la confidencialidad de la información recabada. La aplicación fue grupal, de 20 a 30 personas aproximadamente, en las propias instituciones educativas de pertenencia.

Análisis de datos

Una vez recolectados los datos, en primer lugar, se efectuó un análisis descriptivo y exploratorio para evaluar el comportamiento de los ítems (media, desviación estándar) así como una aproximación descriptiva a la normalidad univariada por medio de la asimetría y curtosis, y el índice estandarizado de asimetría (*Standardized Skew Index*; SSI; Malgady, 2007), esperando valores $< .50$. Asimismo, se valoró el efecto de techo (Terwee et al., 2007), es decir, que más del 15% de los evaluados tengan el puntaje máximo posible en un ítem.

En segundo lugar, se examinó la estructura interna de la EAD mediante el modelamiento exploratorio de ecuaciones estructurales (Asparouhov & Muthén, 2009) con el método de estimación WLSMV orientado a variables ordinales con el programa *Mplus* versión 7 (Muthén & Muthén, 1998-2015).

Fueron evaluados dos modelos en varones y mujeres por separado: primero, tres factores oblicuos bajo un enfoque CFA (M_{1A}) y ESEM (M_{1B}), pues los estudios antecedentes generalmente lo reportan como un modelo razonable, considerando la afinidad teórica de las subescalas y el planteamiento original de sus autores. En segundo lugar, la configuración bifactor (M_2), que modela la elevada correlación entre los factores mediante la especificación adicional de un factor general (FG); este modelo permitiría identificar el monto de varianza común y específica en cada ítem, y valorar la viabilidad de un FG y los tres factores específicos de la EAD. Cabe precisar que fue especificada una rotación geomin ($\epsilon = .05$) para M_{1B} y una rotación

target ortogonal para M_2 (Asparouhov & Muthén, 2009).

En términos generales, los índices de ajuste fueron valorados según puntos de corte ya conocidos: CFI > .90 (Marsh, Hau, & Wen, 2004); límite superior del intervalo de confianza al 90% del RMSEA < .10 (West, Taylor, & Wu, 2012); y WRMR < 1 (DiStefano, Liu, Jiang, & Shi, 2018). De forma complementaria, para el modelo oblicuo ESEM se consideró el índice de simplicidad factorial (ISF; Fleming & Merino, 2005) donde valores > .70 indica que la varianza del ítem es explicada predominantemente por un solo factor, y para el modelo bifactor se recabó evidencia respecto a la presencia del FG mediante diversos indicadores, como el omega jerárquico general (ω_h ; Zinbarg, Yovel, Revelle, & McDonald, 2006), la varianza común explicada (ECV; Sijtsma, 2009) y el porcentaje de correlaciones no contaminadas por la multidimensionalidad (PUC; Reise, Scheines, Widaman, & Haviland, 2013); y con relación a los factores específicos, si sus omega jerárquicos (ω_{hs} ; Zinbarg et al., 2006) son \geq .30 (Smits, Timmerman, Barelds, & Meijer, 2015), se considera que poseen presencia al margen del FG. En este caso, si los valores del ECV y el PUC son superiores a .70, el FG presenta evidencia favorable (Rodríguez, Reise, & Haviland 2016). Los cálculos fueron realizados con el módulo Índices.Bifactor (Dominguez-Lara & Rodríguez, 2017).

Con el modelo con mayor respaldo, se realizó la invarianza de medición entre varones y mujeres, la cual fue evaluada de forma consecutiva y acumulativa en cuanto a diversos parámetros (Pendergast, von der Embse, Kilgus, & Eklund, 2017): la igualdad de configuración total del instrumento (invarianza configuracional), de las cargas factoriales (invarianza métrica) y de los *thresholds* (invarianza fuerte). El grado de invarianza se valoró considerando la variación en la magnitud CFI y RMSEA entre modelos anidados, determinando así que existe evidencia desfavorable para la invarianza de medición si $\Delta CFI < -.01$ y $\Delta RMSEA \geq .01$ (Chen, 2007) o si $\Delta CFI < -.002$ y $\Delta RMSEA \geq .007$ (Meade, Johnson, & Braddy, 2008).

Finalmente, previo análisis de la tau-equivalencia (equivalencia estadística de las cargas factoriales), se reportó la confiabilidad de las puntuaciones mediante el coeficiente α (Cronbach, 1951) con intervalos de confianza (Dominguez-Lara & Merino-Soto, 2015), así como

la confiabilidad del constructo con el coeficiente ω (McDonald, 1999).

Resultados

Análisis descriptivo y exploración inicial de datos

En primera instancia se halló que 17 de los 24 ítems presentaron asimetría moderadas (valores SSI superiores a .25 y menores a .50; [Tabla 1](#)). Del mismo modo, varios ítems presentan el *efecto de techo*.

Tabla 1.

Estadísticos descriptivos de los ítems de la EAD

	<i>M</i>	<i>DE</i>	g_1	g_2	SSI	%Mín	%Máx
Item 1	3.87	.678	-.338	.281	.368	.00	14.7
Item 2	4.04	.646	-.358	.493	.429	.00	21.3
Item 3	3.97	.715	-.344	-.009	.336	.00	21.9
Item 4	3.95	.768	-.955	1.981	.810	1.2	20.5
Item 5	3.86	.638	-.146	.042	.179	.00	13.3
Item 6	4.11	.567	.017	.026	.026	.00	21.9
Item 7	4.10	.701	-.539	.439	.548	.00	27.4
Item 8	3.82	.704	-.379	.223	.382	.00	13.5
Item 9	4.12	.599	-.048	-.208	.067	.00	24.5
Item 10	4.11	.591	-.367	1.198	.525	.00	22.2
Item 11	3.90	.693	-.383	.298	.399	.00	16.4
Item 12	4.14	.689	-.619	1.034	.652	.30	29.7
Item 13	4.05	.661	-.479	1.099	.548	.30	22.8
Item 14	3.95	.672	-.288	.162	.319	.00	18.4
Item 15	3.86	.746	-.221	-.056	.199	.30	18.7
Item 16	4.01	.624	-.225	.298	.289	.90	19.3
Item 17	3.97	.673	-.249	.026	.275	.00	19.6
Item 18	4.08	.645	-.401	1	.482	.30	23.9
Item 19	3.91	.751	-.465	.37	.412	.30	19.9
Item 20	4.14	.605	-.234	.31	.320	.00	25.9
Item 21	3.95	.759	-.801	1.507	.695	.90	21
Item 22	3.70	.841	-.442	-.015	.312	.60	15
Item 23	4.00	.656	-.373	.485	.433	1.7	19.9
Item 24	4.09	.659	-.462	.999	.532	.30	25.1

Nota. *M*: media aritmética; *DE*: desviación estándar; g_1 : asimetría de Fisher; g_2 : curtosis de Fisher. SSI: Índice estandarizado de Asimetría; %Mín: porcentaje de personas con el puntaje mínimo; %Máx: porcentaje de personas con el puntaje máximo.

Evidencias de validez con relación a la estructura interna

Inicialmente se evaluó el ajuste del modelo oblicuo de tres factores (M_1), cuyos índices de ajuste fueron aceptables tanto para varones (CFA: CFI = .954; RMSEA = .073 [IC 90% = .060, .085]; WRMR = .975; ESEM: CFI = .975; RMSEA = .059 [IC 90% = .042, .073]; WRMR = .689) como para mujeres (CFA: CFI = .962; RMSEA = .058 [IC 90% = .049, .067]; WRMR = 1.010; ESEM: CFI = .978; RMSEA = .048 [IC 90% = .037, .058]; WRMR = .731).

Sin embargo, algunos indicadores no favorecen el modelo de tres dimensiones. Para el caso de los CFA, la correlación interfactorial fue elevada en varones ($\phi_{\text{promedio}} = .961$) y mujeres ($\phi_{\text{promedio}} = .893$). Por otro lado, en cuanto al ESEM, el agrupamiento de ítems no es el esperado (Tablas 2 y 3), siendo más bien desordenado y complejo ($ISF < .70$), impidiendo inclusive el cálculo de la confiabilidad del constructo.

Por otro lado, el ajuste del modelo bifactor (M_2) fue aceptable en varones ($CFI = .983$; $RMSEA =$

$.051$ [$IC\ 90\% = .032, .068$]; $WRMR = .577$) y mujeres ($CFI = .986$; $RMSEA = .041$ [$IC\ 90\% = .027, .053$]; $WRMR = .620$), y el FG tuvo evidencia favorable en ambos grupos considerando los índices estadísticos asociados adicho modelo: $ECV (> .80)$, $ECV-I_{\text{promedio}} (> .80)$ y $\omega_h (> .90)$ elevados, mientras que los ω_{hs} fueron muy bajos ($< .20$) (Tablas 2 y 3). De este modo, el análisis de invarianza se realizó teniendo como modelo base el bifactor.

Tabla 2.

Parámetros de los ítems del modelo oblicuo y bifactor de la EAD: mujeres

	CFA			ESEM-oblicuo				ESEM-bifactor				ECV-I	
	F1	F2	F3	F1	F2	F3	ISF	FG	F1	F2	F3		
F1													
Ítem 1	.634			.353	-.063	.454	.526	.584	.307	.124	-.048		.783
Ítem 2	.587			.466	.141	.103	.816	.542	.228	-.157	.053		.850
Ítem 4	.530			.397	-.041	.279	.570	.478	.375	-.025	.018		.619
Ítem 6	.65			.546	.089	.151	.861	.615	.245	-.141	-.015		.863
Ítem 9	.738			.722	-.001	.166	.926	.737	.231	-.187	-.216		.911
Ítem 12	.699			.330	.118	.392	.447	.656	.190	.080	.036		.923
Ítem 14	.806			.325	.137	.502	.542	.729	.300	.120	.126		.855
Ítem 22	.546			-.061	.127	.586	.919	.505	.058	.341	.103		.987
F2													
Ítem 7		.613		.510	.418	-.156	.444	.597	-.018	-.334	.111		.762
Ítem 10		.662		.576	.052	.187	.850	.685	.124	-.123	-.201		.969
Ítem 11		.729		.436	.159	.340	.466	.696	.227	-.031	.055		.998
Ítem 17		.693		.151	.355	.355	.376	.698	-.147	.127	.061		.968
Ítem 18		.743		.281	.331	.307	.094	.791	-.227	.038	-.098		.998
Ítem 20		.703		.464	.520	-.105	.476	.710	-.165	-.301	.105		.848
Ítem 23		.702		.003	.264	.595	.767	.713	-.137	.366	.026		.791
Ítem 24		.719		.114	.067	.707	.949	.718	.083	.367	-.044		.793
F3													
Ítem 3			.677	-.059	.679	.179	.894	.525	.033	-.046	.606		.429
Ítem 5			.594	.332	-.060	.439	.536	.529	.449	.101	.052		.990
Ítem 8			.559	.271	.084	.316	.459	.512	.206	.054	.064		.985
Ítem 13			.740	.185	.447	.269	.490	.686	-.010	.006	.230		.899
Ítem 15			.814	-.018	.625	.374	.645	.685	.059	.094	.525		.630
Ítem 16			.810	.227	.585	.164	.722	.738	-.062	-.093	.300		.858
Ítem 19			.812	.090	.640	.244	.793	.714	-.050	-.012	.407		.755
Ítem 21			.624	.220	.579	-.048	.810	.579	-.186	-.194	.244		.849
F1				1									
F2				.412	1								
F3				.425	.489	1							
ω_h				-	-	-		.917	-	-	-		
ω_{hs}				-	-	-		-	.118	.000	.171		
ECV				-	-	-		.831	-	-	-		

Nota. F1: Eficacia percibida en el ajuste del estudiante; F2: Eficacia percibida en las prácticas instruccionales; F3: Eficacia percibida en el manejo del salón de clase. En negrita: Cargas factoriales de magnitud significativa; en cursiva: Cargas factoriales pertenecientes al factor teórico; FG. Factor general; ω_h : Coeficiente omega jerárquico; ω_{hs} : Coeficiente omega jerárquico de la dimensión; ECV: varianza común explicada; ECV-I: varianza común explicada por ítem

Tabla 3.

Parámetros de los ítems del modelo oblicuo y bifactor de la EAD: varones

	CFA			ESEM-oblicuo				ESEM-bifactor			ECV-I	
	F1	F2	F3	F1	F2	F3	ISF	FG	F1	F2		F3
F1												
Ítem 1	.716			.180	.498	.193	.671	.687	.238	-.049	.145	.893
Ítem 2	.637			.348	.205	.228	.347	.620	.286	.031	-.131	.825
Ítem 4	.689			.063	.419	.357	.499	.648	.247	.104	.177	.873
Ítem 6	.763			.409	.392	.130	.426	.756	.115	-.104	-.014	.977
Ítem 9	.837			.380	.354	.291	.149	.821	.116	.013	.018	.980
Ítem 12	.736			.381	.161	.364	.374	.726	.099	.123	-.074	.982
Ítem 14	.759			.407	.570	-.050	.566	.766	-.018	-.268	.113	.999
Ítem 22	.619			.196	.417	.143	.624	.606	.053	-.058	.138	.992
F2												
Ítem 7		.694		-.002	.020	.813	.999	.623	.127	.508	.087	.601
Ítem 10		.844		.338	.355	.336	.035	.813	.341	.048	-.009	.997
Ítem 11		.757		.819	-.006	.107	.975	.776	-.109	-.079	-.339	.990
Ítem 17		.682		.543	.192	.099	.798	.705	-.292	-.102	-.074	.979
Ítem 18		.784		.277	.400	.284	.256	.786	-.303	.026	.230	.999
Ítem 20		.718		.112	-.136	.890	.943	.661	-.116	.621	.020	.531
Ítem 23		.784		.655	-.026	.333	.713	.794	-.133	.100	-.249	.984
Ítem 24		.787		.367	.293	.302	.149	.774	.004	.048	.025	.996
F3												
Ítem 3			.630	.005	.427	.331	.544	.588	.222	.081	.222	.875
Ítem 5			.716	.247	.658	-.038	.818	.698	.309	-.277	.152	.955
Ítem 8			.715	.668	-.014	.221	.856	.726	.107	.010	-.366	.797
Ítem 13			.856	.193	.550	.308	.556	.836	.025	-.003	.249	.919
Ítem 15			.800	-.051	.584	.438	.549	.748	.139	.124	.359	.813
Ítem 16			.877	.523	.337	.213	.469	.890	-.202	-.066	.000	1.000
Ítem 19			.683	-.128	.567	.386	.544	.638	-.137	.108	.541	.582
Ítem 21			.593	.107	.000	.610	.956	.548	.128	.374	-.003	1.000
F1				1								
F2				.480	1							
F3				.461	.485	1						
ω_h				-	-	-		.956	-	-	-	
ω_{hs}				-	-	-		-	.035	.035	.036	
ECV				-	-	-		.887	-	-	-	

Nota. F1: Eficacia percibida en el ajuste del estudiante; F2: Eficacia percibida en las prácticas instruccionales; F3: Eficacia percibida en el manejo del salón de clase. En negrita: Cargas factoriales de magnitud significativa; en cursiva: Cargas factoriales pertenecientes al factor teórico; FG: Factor general; ω_h : Coeficiente omega jerárquico; ω_{hs} : Coeficiente omega jerárquico de la dimensión; ECV: varianza común explicada; ECV-I: varianza común explicada por ítem.

Tabla 4.

Invarianza de medición del modelo bifactor según el sexo

	$\chi^2(g)$	CFI	RMSEA (IC90%)	WRMR	Δ CFI	Δ RMSEA
Configural	501.436 (372)**	.983	.045 (.034, .055)	.850		
Métrica	602.171 (452)**	.981	.044 (.034, .053)	1.107	-.002	-.001
Fuerte	646.437 (518)**	.983	.038 (.027, .047)	1.156	.002	-.006

Invarianza de medición

En vista de que el ajuste del modelo bifactor es aceptable en ambos grupos, se evaluó la invarianza de medición, obteniendo resultados

favorables en cuanto a la invarianza configural, métrica y fuerte (Tabla 4). Es decir, el modelo que resalta la presencia de un FG en lugar de tres específicos tiene un funcionamiento similar en

varones y mujeres.

Evidencias de Consistencia Interna

Previo al análisis de la confiabilidad de las puntuaciones con la muestra total por medio del coeficiente α se evaluó la tau-equivalencia. En primer lugar, el modelo unidimensional (congenérico) obtuvo índices de ajuste aceptables (CFI = .949; RMSEA = .065 [IC 90% = .059, .072]; WRMR = 1.191), y una vez que se especificó la igualdad estadística de las 24 cargas factoriales, obtuvo magnitudes similares al congenérico (CFI = .931; RMSEA = .073 [IC 90% = .067, .079]; WRMR = 1.719). Entonces, siendo las diferencias entre modelos poco significativa (Δ CFI = -.018; Δ RMSEA = .008), se concluye a favor de la tau-equivalencia, siendo el coeficiente α de magnitud elevada (.933, [IC 95% .918, .945]). En cuanto a la confiabilidad del constructo se calculó el coeficiente ω , obteniendo un valor también elevado (.954).

Discusión

El presente estudio tuvo como objetivo evaluar un aspecto fundamental de la validez de las inferencias respecto la EAD, es decir, su dimensionalidad o características estructurales. Sin estos resultados, el resto de las evidencias de validez serían de limitada utilidad. Cabe resaltar que en los antecedentes se enfocaron parcialmente en el aspecto estructural del EAD, es decir, en los modelos oblicuos, pero se dejó de explorar aquellas características estructurales relevantes que permiten comprender la relación de los ítems con los constructos que evalúa la escala (similitud de las cargas factoriales entre los ítems, la relación de los ítems con las otras dimensiones y su funcionamiento diferencial de acuerdo al sexo). Asimismo, a pesar de las elevadas correlaciones interfactoriales, en trabajos posteriores al original no fue evaluado un constructo de orden superior (Tschannen-Moran & Woolfolk Hoy, 2001), y aunque las autoras hallaron un factor de segundo orden, no fue tomado en cuenta por los investigadores más adelante. En tal sentido, los resultados presentados en este reporte constituyen un avance en relación a las estrategias analíticas para la EAD en docentes peruanos.

Los resultados obtenidos sobre la unidimensionalidad del EAD no parecen inesperados debido a que se ha demostrado en reportes psicométricos que el comportamiento de

los ítems varía cuando la EAD es analizada a nivel de estructura interna en profesionales (Chang & Engelhard, 2016; Fives & Buehl, 2010), en universitarios que realizan prácticas profesionales (Duffin et al., 2012; Fives & Buehl, 2010; Khairani & Razak, 2012) o con ambas muestras de forma conjunta (Tschannen-Moran & Woolfolk Hoy, 2001). Por ejemplo, cuando el análisis se realizó en profesionales, las dimensiones tuvieron un mayor grado de diferenciación empírica, es decir, sus correlaciones interfactoriales no fueron tan problemáticas ni se evidenció una complejidad factorial marcada. En cambio, cuando los análisis involucraban estudiantes en prácticas profesionales, o las dos muestras de forma conjunta, la superposición (o correlaciones interfactoriales muy elevadas) entre los factores originales fue más evidente.

Esto indicaría que las creencias de autoeficacia se logran diferenciar a medida que el profesional adquiere mayores competencias y es capaz de distinguir entre aquellas vinculadas al contenido de la clase (*Eficacia percibida en las prácticas instruccionales*) y las que conciernen a la relación docente-estudiante (*Eficacia percibida en el ajuste del estudiante; Eficacia percibida en el manejo del salón de clase*). Los estudiantes practicantes no diferencian a nivel cognitivo estos aspectos, considerando de forma holística la dinámica de clase y esta situación podría deberse en la falta de experiencia dentro del salón de clases (Tschannen-Moran & Woolfolk Hoy, 2001).

Entonces, a diferencia de los estudios precedentes que defendían el argumento de una escala multidimensional, la identificación de una sola dimensión de la EAD subyacente a todos los ítems se sugiere, al menos en esta muestra de docentes peruanos de entidades públicas, que todos los componentes que anteriormente parecían diferenciados pueden estar altamente relacionados con un único constructo, el mismo cuyas conductas estarían influenciadas mutuamente en las tareas de rol que el docente enfrenta en el aula de clase. Entonces, los docentes participantes no distinguen entre las tres dimensiones de la EAD, respondiendo sobre sus creencias de eficacia docente desde una perspectiva general probablemente porque prefieren autoevaluar su actuación en el aula desde una perspectiva integrativa. Ello se podría corroborar en otras investigaciones con variables ligadas al desempeño docente.

Cuando la escala fue creada, Tschannen-Moran y Woolfolk Hoy (2001) sostuvieron que el modelo psicométrico de la EAD se encontraba en fase exploratoria, por lo tanto, la posibilidad de encontrar resultados diferenciados sobre su estructura latente (como la unidimensionalidad del constructo), es viable. Por tal motivo, el EAD podría presentar variantes en cuanto a la consideración de su dimensionalidad tomando en cuenta las características idiosincráticas de la muestra evaluada.

Los resultados encontrados en el presente estudio respecto a la estructura interna de la EAD en maestros peruanos son iniciales y dan apertura a la replicabilidad de resultados para constatar la consistencia de la unidimensionalidad de la EAD. Los instrumentos de medición de la autoeficacia del docente precisamente se han caracterizado por ser unidimensionales (Schwarzer, Schmitz, & Daytner, 1999) y multidimensionales (Skaalvik & Skaalvik, 2007), siendo estos formatos de evaluación aceptados pues siguen una línea teórica que se ajusta a la evaluación de contextos específicos de la autoeficacia (Bong, 2006).

Aunque se constató la presencia de un FG más allá de los específicos, algunas de estas conductas parecen aportar aspectos únicos y no comunes al resto de conductas muestreadas por los ítems. Como estos ítems aportan varianza más allá del error de muestreo y de medición, y contienen sustancial varianza común y específica, entonces pueden aportar a la descripción diferencial de la percepción de autoeficacia. Entonces, se podrían elegir estos ítems que contienen ambas fuentes de varianza (común y específica) para representar un contenido parsimonioso y no redundante, y desarrollar una versión abreviada de la EAD. Con todo, aunque existen versiones abreviadas de la EAD, estas no han sido validadas en los participantes de esta investigación, y por lo tanto existe una cuestión abierta al respecto. Esta potencial versión breve debe ser evaluada con evidencias de validez estructural y de correlatos con otros constructos a fin de brindarle mayor fortaleza a las interpretaciones posteriores.

Un aspecto no abordado en los estudios anteriores es la invarianza entre varones y mujeres. Si bien el sexo no fue una variable abordada anteriormente, la elevada proporción de mujeres en las muestras estudiadas (Duffin et al., 2012; Fernández & Merino, 2012; Fives & Buehl, 2010; Tschannen-Moran & Woolfolk Hoy, 2001) podría

hacer pensar una estructura sesgada hacia dicho sexo, pero las evidencias halladas indican equivalencia estructural respecto a las creencias de autoeficacia según el sexo, es decir, que las creencias de *autoeficacia docente* se evalúan de forma equitativa entre varones y mujeres, lo que permitiría más adelante llevar a cabo estudios comparativos según el sexo.

Finalmente, respecto a la estimación de la confiabilidad de las puntuaciones y constructo, las magnitudes de los coeficientes de confiabilidad encontrados son alentadoras.

Los resultados obtenidos no solo dan una nueva perspectiva sobre cómo conceptualizar la percepción de autoeficacia en el tipo de muestra examinada, sino también ponen en relieve que la presunción de equivalencia psicométrica no es necesariamente viable entre culturas, hasta ahora no demostrado empíricamente con la EAD.

Entre las limitaciones se puede destacar que la muestra de docentes utilizada en el estudio pertenece sólo a escuelas públicas, no habiéndose considerado escuelas privadas que, en la ciudad de Lima, son similares en cantidad, pero no en calidad educativa, lo cual podría estar ocasionado algún sesgo en la interpretación de resultados. Por tal motivo, sería importante realizar otros estudios que consideren instituciones públicas y privadas y que las muestras sean más amplias para poder tener resultados más aproximados a la realidad del colectivo docente.

Pese a la importancia del rol del profesorado en la educación en Perú, aún son escasos los estudios que se enfocan en la autoeficacia docente, por lo que sería conveniente continuar esta línea de investigación incorporando aquellas variables psicológicas relevantes a dicha labor, así como aquellas vinculadas a la vivencia en aula (e.g., años de experiencia, nivel de enseñanza, grado de especialización, etc.), ya que pueden ser aspectos sustanciales que podrían tener relevancia práctica y entender de una forma más contextual las creencias de autoeficacia que posee el docente peruano.

Referencias

- Alessandri, G., Borgogni, L., Schaufeli, W. B., Caprara, G. V., & Consiglio, C. (2015). From positive orientation to job performance: The role of work engagement and self-efficacy beliefs. *Journal of Happiness Studies*, 16(3), 767–788. doi: 10.1007/s10902-014-9533-4.

- Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059. doi: 10.6018/analesps.29.3.178511
- Asociación Médica Mundial (1964). *Declaración de Helsinki*. Helsinki: AMM. Recuperado de: http://www.conamed.gob.mx/prof_salud/pdf/helsinki.pdf
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16(3), 397-438. doi: 10.1080/10705510903008204
- Bandura, A. (1986). *Social foundations of thoughts and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Blanco, H., Ornelas, M., Aguirre, J. F., & Guedea, J. C. (2012). Autoeficacia percibida en conductas académicas. Diferencias entre hombres y mujeres. *Revista Mexicana de Investigación Educativa*, 17(53), 55-59.
- Bong, M. (2006). Asking the right question. How confident are you that you could successfully perform these tasks? En F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 287–305). Greenwich, CT: Information Age Publishing
- Calderón-De la Cruz, G., Domínguez-Lara, S., & Arroyo-Rodríguez, F. (2018). Análisis psicométrico preliminar de una medida breve de autoeficacia profesional en trabajadores peruanos: AU-10. *Psicogente* 21(39), 12-24. doi: 10.17081/psico.21.39.2819
- Chang, M. L., & Engelhard, G. (2016). Examining the teachers' Sense of Efficacy Scale at the Item Level with Rasch Measurement Model. *Journal of Psychoeducational Assessment*, 34(2), 177-191, doi: 10.1177/0734282915593835
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504. doi: 10.1080/10705510701301834
- Colegio de Psicólogos del Perú (2017). *Código de ética y deontología*. Recuperado de: https://www.cpsp.pe/documentos/marco_legal/codigo_de_etica_y_deontologia.pdf
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. doi: 10.1007/BF02310555
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. doi: 10.1207/S15327965PLI1104_01.
- Díaz, H. (2015). *Formación docente en el Perú. Realidad y tendencias*. Lima: Fundación Santillana.
- DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness? *Structural Equation Modeling*, 25(3), 453-466. doi: 10.1080/10705511.2017.1390394
- Dominguez-Lara, S., & Merino-Soto, C. (2015). ¿Por qué es importante reportar los intervalos de confianza del coeficiente alfa de Cronbach? *Revista Latinoamericana de Ciencias Sociales, niñez y Juventud*, 13(2), 1326-1328.
- Dominguez-Lara, S., & Rodriguez, A. (2017). Índices estadísticos de modelos bifactor. *Interacciones*, 3(2), 59-65. doi: 10.24016/2017.v3n2.51
- Duffin, L., French, B., & Patrick, H. (2012). The Teacher's Sense of Efficacy Scale: confirming the factor structure with beginning pre-service teachers. *Teaching and Teacher Education*, 28(6), 827 – 834. doi: 10.1016/j.tate.2012.03.004
- Fernández, M. (2008). Desempeño docente y su relación con orientación a la meta, estrategias de aprendizaje y autoeficacia: un estudio con maestros de Lima, Perú. *Universitas Psychologica*, 7(2), 385-401.
- Fernández, M., & Merino, C. (2012). Resultados psicométricos preliminares de la Escala de Autoeficacia Percibida en maestros de Lima. *Psicogente*, 15(28), 314 – 322.
- Fives, H., & Buehl, M. (2010). Examining the factor structure of the Teacher's Sense of Efficacy Scale. *The Journal of Experimental Education*, 78(1), 118-134. doi: 10.1080/00220970903224461
- Fleming, J., & Merino, C. (2005). Medidas de simplicidad y ajuste factorial: Un enfoque para la construcción y revisión de escalas derivadas factorialmente. *Revista de Psicología*, 23(2), 252-266.
- Graham, J. M. (2006). Congeneric and essentially tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*, 66(6), 930– 944. doi: 10.1177/0013164406288165
- Khairani, A. Z., & Razak, N. (2012). An Analysis of the Teacher's Sense Efficacy Scale within the Malasyan Context using the Rasch Measurement Model. *Procedia – Social and Behavioral Sciences*, 69, 2137-2142. doi: 10.1016/j.sbspro.2012.12.178
- Malgady, R. (2007). How skew are psychological data? A standardized index of effect size. *The Journal of General Psychology*, 134(3), 355-359. doi: 10.3200/GENP.134.3.355-360
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320-341. doi: 10.1207/s15328007sem1103_2
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied*

- Psychology*, 93(3), 568 – 592. doi: 10.1037/0021-9010.93.3.568
- Mielniczuk, E., & Laguna, M. (2018). Positive Affect Mediates the Relationship Between Self-efficacy and Innovative Behavior in Interpreneurs. *The Journal of Creative Behavior*, 1-12. doi: 10.1002/jocb.364
- Ministerio de Educación (2010). *Resultados de la Evaluación PISA 2009*. Lima: Ministerio de Educación. Recuperado de: <http://umc.minedu.gob.pe/wp-content/uploads/2010/06/PISA2009-versi%C3%B3n-completa-para-secundaria.pdf>
- Mojavezi, A., & Poodineh, M. (2012). The Impact of Teacher Self-efficacy on the students' Motivation and Achievement. *Theory and Practice in Language Studies*, 2(3), 483-491. doi: 10.4304/tpis.2.3.483-491
- Muthén, L. K., & Muthén, B. O. (1998 - 2015). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology*, 60, 65-82. doi: 10.1016/j.jsp.2016.11.002
- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73(1), 5 - 26. doi: 10.1177/0013164412449831
- Rivero, J. (2007). *Educación docente y clase política en el Perú*. Lima: Tarea.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137 - 150. doi: 10.1037/met0000045
- Schwarzer, R., Schmitz, G. S., & Daytner, G. T. (1999). *The Teacher Self-Efficacy scale* [On-line publication]. Recuperado de: http://www.userpage.fu-berlin.de/~health/teacher_se.htm
- Scholz, H., Gutiérrez, B., & Schwarzer, R. (2002). Is General Self-Efficacy a universal construct? Psychometric Findings from 25 countries. *European Journal Psychological Assessment*, 18(3), 242-251. doi: 10.1027/1015-5759.18.3.242
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120. doi: 10.1007/s11336-008-9101-0
- Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, 99(3), 611-625.
- Smits, I. A. M., Timmerman, M. E., Barelds, D. P. H., & Meijer, R. R. (2015). The Dutch symptom checklist-90-revised: is the use of the subscales justified? *European Journal of Psychological Assessment*, 31(4), 263-271. doi: 10.1027/1015-5759/a000233
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783 – 805. doi: 10.1016/S0742-051X(01)00036-1
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 2(68), 202-248. doi: 10.3102/00346543068002202
- Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality Criteria Were Proposed for Measurement Properties of Health Status Questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34-42. doi: 10.1016/j.jclinepi.2006.03.012
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, UNESCO (2017). *Revisión de las políticas públicas del sector educación en el Perú*. Francia: UNESCO. Recuperado de: <https://unesdoc.unesco.org/ark:/48223/pf0000260875>
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. En R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209-234). New York, NY: Guilford Press
- Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for wh. *Applied Psychological Measurement*, 30(2), 121-144. doi: 10.1177/0146621605278814