



Evidencias de validez de la Escala de Sentimiento de Autoeficacia Docente en una muestra española

José-Antonio Bueno-Álvarez*, Margarita-Carmen Martín-Martín, Enrique Navarro-Asencio, and Isabel-Inmaculada Asensio-Muñoz

Universidad Complutense de Madrid (Spain)

Resumen: El sentimiento de autoeficacia docente (SAD) es un constructo de compleja medida pero relevante por su relación con la calidad de la educación. Con este trabajo se pretende acumular evidencias de consistencia y validez para su uso en España de una versión del Teachers' Sense of Efficacy Scale (TSES). Los datos proceden de un colectivo español de futuros profesores de infantil, primaria y secundaria ($N = 744$) y se analizan desde un enfoque multivariado adecuado para escala ordinal, mediante análisis factorial confirmatorio (ULSMV) y árboles de decisión (CHAID y CART). Tras evaluar tres modelos de medida, se aportan evidencias de validez de una estructura del constructo en tres factores y 17 ítems, con índices de ajuste aceptables. Además, se avala tanto la convergencia y consistencia del constructo, como la utilidad de los ítems para la predicción de la autoeficacia docente global en los colectivos de estudiantes de grado y de máster que componen la muestra estudiada.

Palabras clave: Adaptación. Análisis factorial confirmatorio. Árboles de decisión. Autoeficacia docente. Profesores en formación. Validación.

Title: Evidence for the validity of the Teachers' Sense of Self-Efficacy Scale in a Spanish sample.

Abstract: Teachers' sense of self-efficacy is a hard construct to measure but is important in view of its relationship with the quality of education. This work aims to gather evidence regarding the consistency and validity of a version of the Teachers' Sense of Efficacy Scale (TSES) for use in Spain. The data come from a Spanish group of future early-years, primary and high school teachers ($N = 744$), and they are analysed using a multivariate approach suitable for ordinal scales, with confirmatory factor analysis (ULSMV) and decision trees (CHAID and CART). After evaluating three measurement models, evidence for the validity of a construct structure with three factors and 17 items with acceptable fit indices is provided. In addition, the convergence and consistency of the construct are both endorsed, as is the usefulness of the items for predicting overall teacher self-efficacy in the groups of undergraduate and master's students in the sample studied.

Keywords: Adaptation. Confirmatory factor analysis. Decision trees. Teaching self-efficacy. Teachers in training. Validation.

Introducción

El sentimiento de autoeficacia docente (SAD) es una de las cualidades que más y mejor caracteriza al profesor en su desempeño educativo, al ser un indicador importante del grado de satisfacción profesional del propio docente y un buen predictor del rendimiento que pueden alcanzar sus alumnos (Chesnut y Burley, 2015; Zee y Koomen, 2016), de ahí el interés que suscita como tópico de investigación en Psicología de la Educación.

Como constructo, se enmarca en la teoría del aprendizaje social propuesta por Bandura (1977), que permite explicar el comportamiento del ser humano en múltiples contextos (Azzi & Polydoro, 2006; Salanova et al., 2004).

El SAD se define como la creencia que el docente mantiene acerca de sus capacidades para enfrentarse a las tareas cotidianas que conlleva el proceso de enseñanza-aprendizaje, a nivel individual y colectivo. En tanto que creencia, es un constructo complejo, con componentes cognitivos y afectivos, que marca el modo en que el individuo interpreta su realidad profesional y se implica con ella (Asensio-Muñoz & Ruiz de Miguel, 2017), por lo que actúa como mediador entre el conocimiento y la acción docente (Pajares, 1992). La complejidad aludida se incrementa al tratarse de un senti-

miento afectado también por el contexto donde se desarrolla la acción docente (Friedman & Kass, 2002).

Tschannen-Moran et al. (1998; Tschannen-Moran y Woolfolk-Hoy, 2001) presentan el modelo teórico de SAD que más aceptación tiene en la actualidad, por su valor explicativo y predictivo. En él, el SAD se caracteriza por dos componentes que se influyen mutuamente y determinan, tanto a nivel conductual como cognitivo, el quehacer cotidiano del profesor: el análisis de la tarea y la autoevaluación de la competencia personal (Tschannen-Moran & McMaster, 2009).

El SAD se ve afectado no solo por las tareas docentes intra-clase, sino también por el contexto escolar y organizativo del centro: el docente funciona simultáneamente como líder del salón de clase, a nivel de tarea, y como empleado del centro, a nivel de relación (Tschannen-Moran et al., 2015).

Para evaluar esta propuesta, Tschannen-Moran y Woolfolk-Hoy (2001), desarrollan el Ohio State Teacher Efficacy Scale (OSTES), que tras diversos análisis da lugar al Teachers' Sense of Efficacy Scale (TSES), de 24 ítems (con una versión reducida de 12), agrupados en tres factores: eficacia para la implementación de estrategias instruccionales (8 ítems); eficacia para la gestión del aula (8 ítems); y eficacia para la implicación de los estudiantes (8 ítems). En la actualidad es el instrumento más empleado para la evaluación de este constructo a nivel internacional (Kleinsasser, 2014) y en la investigación iberoamericana (Bueno-Álvarez et al., 2022).

En el contexto hispanoparlante, en el que nos centramos en este artículo, se reportan aplicaciones del TSES por parte de Bermejo y Prieto (2005), Perandones y Castejón (2007) y Rodríguez et al. (2009), pero la primera versión que se valida

* Correspondence address [Dirección para correspondencia]:

José Antonio Bueno Álvarez. Universidad Complutense de Madrid. Departamento de Investigación y Psicología en Educación. Facultad de Educación C.F.P. C/ Rector Royo Villanova s/n. 28040. Madrid (España).

E-mail: alvarez@ucm.es

(Artículo recibido: 26-04-2022; revisado: 16-09-2022; aceptado: 10-10-2022).

es la de Covarrubias (2014) y Covarrubias y Mendoza (2016). Estas autoras reducen el número de ítems a 17 e identifican un factor más, proponiendo así un instrumento con una estructura factorial compuesta por cuatro componentes. Otras investigaciones llevadas a cabo con posterioridad a Covarrubias (2014), o no realizan estudio psicométrico alguno (Mar-

tínez-Luque et al., 2017; Perandones & Herrera, 2017; Suárez-Escobar, 2018; Uriarte et al., 2019), o reproducen la misma estructura factorial original (Burgueño et al., 2019; Domínguez-Lara et al., 2019; Lera et al., 2021; Salas-Rodríguez et al., 2021). Una descripción resumida de estos estudios se ofrece en la Tabla 1.

Tabla 1*Aplicaciones del TSES en español.*

Autores	Muestra	Forma	Estructura	Validación
Bermejo y Prieto (2005)	71 docentes de enseñanza secundaria obligatoria españoles	TSES en español (no se especifica más) con escala de respuesta de 9 puntos	3 subescalas, 12 ítems	
Burgueño et al. (2019)	358 estudiantes máster de formación de profesorado de secundaria españoles	TSES-24, TSES-12 y TSES-11. Traducción inversa con escala de respuesta de 9 puntos	3 subescalas, 24, 12 y 11 ítems respectivamente	Validez: Análisis Factorial Exploratorio (AFE); Análisis Factorial Confirmatorio (AFC): máx verosimilitud, bootstrapping. Bondad ajuste: χ^2 , ratio entre chi cuadrado y los grados de libertad (χ^2 / gl), CFI, IFI, TLI, RMSEA, SRMR.
Covarrubias (2014); Covarrubias y Mendoza (2016)	544 docentes en activo chilenos	Traducción inversa TSES con escala de respuesta de 5 puntos	4 subescalas, 17 ítems	Constructo: AFE; AFC: máx verosimilitud, bootstrapping. Bondad ajuste: χ^2 , ratio entre chi cuadrado y los grados de libertad (χ^2 / gl), CFI, IFI, RMSEA, SRMR. Fiabilidad: alpha de Cronbach
Domínguez-Lara et al. (2019)	347 docentes educación no superior peruanos	TSES en español (EAD) con escala de respuesta de 5 puntos	3 subescalas, 24 ítems	Validez: Ecuaciones estructurales (WLSMV) CFI, RMSEA, ESEM:ISF. Fiabilidad: alpha de Cronbach
Lera et al. (2021)	2242 niñas y niños de dos comunidades españolas	Traducción (sin informar de cómo se hizo) y adaptación del OSTES para que lo puedan responder los niños en una escala de 5 puntos	3 subescalas, 24 ítems	AFE AFC ordinal MAFC Fiabilidad: omega.
Martínez-Luque et al. (2017)	52 docentes universitarios españoles	TSES en español sin informar de cómo se hizo la traducción con escala de respuesta de 9 puntos	3 subescalas, 24 ítems	No se calcula, pero se reporta el alpha de otro estudio
Perandones y Castejón (2007); Perandones y Herrera (2017)	564 docentes EI, EP dominicanos	TSES-short form. Traducción directa de Perandones y Castejón (2017) con escala de respuesta de 9 puntos	3 subescalas, 24 ítems	Validez de constructo: AFE componentes principales y varimax, KMS y Bartlett. Fiabilidad: alpha de Cronbach
Rodríguez et al. (2009)	95 docentes universitarios españoles	TSES en español (no especifican más)	3 subescalas	
Salas-Rodríguez et al. (2021)	190 docentes de EP y ES mexicanos	TSES-12. Traducción inversa con escala de respuesta de 9 puntos	3 subescalas	Constructo: AFC: máx verosimilitud. Bondad ajuste: χ^2 , ratio entre chi cuadrado y los grados de libertad (χ^2 / gl), CFI, TLI, RMSEA, SRMR. Fiabilidad: alpha de Cronbach
Suárez-Escobar (2018)	86 docentes educación no superior colombianos	TSES español sin informar de cómo se hizo la traducción con escala de respuesta de 9 puntos	3 subescalas, 24 ítems	Fiabilidad: alpha de Cronbach
Uriarte et al. (2019)	300 maestros dominicanos	TSES español sin informar de cómo se hizo la traducción con escala de respuesta de 9 puntos	3 subescalas, 24 ítems	Fiabilidad: alpha de Cronbach

Vista la utilización que se está haciendo del TSES en español, la diversidad de traducciones que existen y tratándose de un constructo posiblemente todavía elusivo (Tschannen-

Moran & Woolfolk-Hoy, 2001), parece de interés seguir acumulando evidencias acerca de la validez estructural a la que llegan Covarrubias y Mendoza (2016) con profesores

chilenos. En el presente estudio, se trabaja con una muestra de profesores españoles en formación y con una metodología de análisis factorial confirmatorio (AFC) ordinal, más adecuada normalmente a los datos procedentes de escalas tipo Likert. También se pretende avanzar en el estudio de la validez concurrente de la escala mediante la técnica de árboles de decisión, igualmente adecuadas con datos ordinales (Blanco et al., 2017). En consecuencia, a continuación, se presenta un estudio de validación, entendido como un proceso de acumulación de evidencias acerca de una medida para el propósito que esta persigue, en el que son precisamente las características de los datos y los objetivos de la medida los que determinan el tipo de aproximación metodológica más conveniente en cada caso (AERA, APA y NCME, 2014). El objetivo del trabajo es ahondar en la definición operativa del constructo en español y en la validez de la medida propuesta por Covarrubias y Mendoza (2016), reuniendo nuevas evidencias que apoyen 1) la dimensionalidad de la autoeficacia docente, así como la convergencia y consistencia interna del modelo de medida propuesto y las posibles diferencias entre los estudiantes de grado y máster; y 2) la utilidad de los ítems del TSES adaptado para la identificación de los ítems que mejor predicen la percepción que los estudiantes tienen de su autoeficacia docente global, teniendo en consideración si estos son de grado o de máster.

Método

Participantes

Los datos corresponden a una muestra incidental de 744 estudiantes de la Universidad Complutense de Madrid que cursan primer año en los grados de Maestro en Educación Primaria (EP) (28%) y Educación Infantil (EI) (19.7%), en los dobles grados de EP con EI o con Pedagogía (18.2%) y el Máster de Formación del Profesorado de Secundaria (MFPS) (34%) con diferentes especialidades. La edad media de 21.13 años ($DT = 4.573$). El porcentaje de mujeres es del 73.8%.

Instrumento

Se aplicó la versión de Covarrubias (2014) del Teachers' Sense of Efficacy Scale – TSES, que adaptan el "Teachers Self Efficacy Scale" (TSES) de Tschannen-Moran y Hoy (2001). Covarrubias (2014) hace una traducción inversa del instrumento original. En la adaptación al español se elige una escala de respuesta de cinco puntos y se aplica a una muestra de profesores chilenos en activo. Los análisis factoriales, exploratorios y confirmatorios, del instrumento adaptado, publicados por Covarrubias y Mendoza (2016), avalan, en la muestra hispanohablante utilizada, la adecuación de una medida con cuatro factores y 17 ítems. Covarrubias (2014) y Covarrubias y Mendoza (2016) informan de la estructura que emerge en la muestra chilena, que es la siguiente:

- A. Eficacia en la implicación de los estudiantes. Factor relacionado con la capacidad de los profesores para motivar e implicar a los estudiantes en la enseñanza y en el aprendizaje (A1, A2, A3, A4, que, como se muestra en la tabla 3, equivalen a los ítems de nuestro estudio 70, 71, 72 y 73 y a los ítems de la escala original 24, 22, 19 y 23)
- B. Eficacia en las estrategias de enseñanza y aprendizaje (EA). Factor enfocado a la capacidad de los profesores para utilizar diferentes estrategias y/o metodologías para enseñar y suscitar el aprendizaje en los estudiantes. (B1, B2, B3, B4, que, como se muestra en la tabla 3, equivalen a los ítems de nuestro estudio 78, 75, 76 y 77 y a los ítems de la escala original 5, 7, 3 y 2)
- C. Eficacia en el manejo de la clase. Factor enlazado con la capacidad de los profesores para manejar o gestionar la conducta, la disciplina y el orden favoreciendo un clima positivo de aula. (C1, C2, C3, C4 y C5, que, como se muestra en la tabla 3, equivalen a los ítems de nuestro estudio 74, 85, 79, 81 y 83 y a los ítems de la escala original 9, 10, 11, 13 y 14)
- D. Eficacia en la atención a la singularidad de los estudiantes. Factor conectado con la capacidad de los profesores para adaptar la enseñanza a las necesidades y/o demandas específicas de aprendizaje en cada estudiante. (D1, D2, D3, D4, que, como se muestra en la tabla 3, equivalen a los ítems de nuestro estudio 82, 84, 80 y 86 y a los ítems de la escala original 6, 1, 4 y 8)

Dicha versión en español, reducida a 17 ítems, es la que se utiliza en el presente estudio.

Procedimiento

El instrumento fue administrado de manera presencial por aquellos docentes de grupos que manifestaron su voluntad de colaborar en la investigación. Una vez dadas unas instrucciones generales, los participantes respondieron con consentimiento informado y de manera voluntaria al cuestionario de autoeficacia docente, junto con otras cuestiones sociodemográficas y de alta inferencia relevantes para el estudio.

Análisis estadístico y psicométrico

Para reunir evidencias en torno a la estructura del constructo, y dar respuesta al primer objetivo, tras la comprobación de casos perdidos, de falta de normalidad multivariada mediante el estadístico de curtosis multivariada de Mardia, mediante el estadístico b2 (Cain et al., 2017) y la identificación de casos extremos a través de la estimación de las distancias de Mahalanobis, se calculan los valores de los parámetros de los modelos confirmatorios a partir de una matriz de correlaciones policóricas y el método de estimación de mínimos cuadrados no ponderados robustos (ULSMV) (Li, 2014; Xia, 2016).

Debido al carácter ordinal de los ítems, la función de en-

lace es un modelo probit que estima interceptos y umbrales. Estos umbrales señalan el nivel o puntuación necesaria en la dimensión para situarse en las distintas categorías de respuestas, es decir, el primer paso muestra el nivel para indicar la categoría 1 respecto a la 0, el segundo paso la categoría 2 respecto a la 1, y así sucesivamente.

Se han probado tres modelos de medida. El primero replica la estructura original de tres dimensiones de Tschanen-Moran y Hoy (2001), el segundo es el modelo de cuatro factores de Covarrubias (2014) y Covarrubias y Mendoza (2016) y, el tercero, es el modelo definitivo resultado de este estudio. La composición de ítems y dimensiones se presenta en la Tabla 2.

Tabla 2

Modelos de medida estimados (ítems y factores latentes).

	Original	Covarrubias	Modelo propuesto
Estrategias EA	75	75	75
	76	76	76
	77	77	77
	78	78	78
	80		
	82		82
	84		
	86		86
Gestión del aula	74	74	74
	79	79	79
	81	81	81
	83	83	83
	85	85	85
Implicación de estudiantes	70	70	
	71	71	71
	72	72	72
	73	73	73
			80
		84	
Eficacia en la atención a singularidad de los estudiantes		80	
		82	
		84	
		86	

Las variaciones principales respecto al original tienen que ver con la cuarta dimensión de la estructura de Covarrubias (2014) y Covarrubias y Mendoza (2016), centrada en la atención a las particularidades de los estudiantes utilizando estrategias alternativas de enseñanza (ítem 80), ajustando el nivel (ítem 82), utilizando estrategias de evaluación distintas a las habituales (ítem 84) y ofreciendo desafíos adecuados a los más capaces (ítem 86). En el modelo original de tres dimensiones estos cuatro ítems se encuentran en el factor de eficacia de estrategias de enseñanza-aprendizaje (EA).

La utilización de estrategias de enseñanza y la evaluación alternativa (ítems 80 y 84) en el modelo final propuesto se sitúan en la dimensión de implicación de los estudiantes, junto con el fomento del pensamiento crítico (ítem 71), la motivación a los estudiantes menos interesados (ítem 72) y el fomento de la creatividad (ítem 73).

En el factor de implicación de estudiantes, los tres mode-

los incluyen los ítems 71, 72 y 74. Además, tanto en el modelo original como en el de Covarrubias (2014), en este factor se incluye un ítem relativo a la comunicación con los estudiantes más difíciles (ítem 70). En cambio, en el modelo final, este ítem se sitúa en el factor sobre la eficacia en la gestión del aula, junto con el control del mal comportamiento en clase (ítem 74), el manejo de los estudiantes con mal comportamiento (ítem 79), el evitar que los estudiantes problemáticos interrumpen la clase (81), el manejo de los estudiantes que se enfrentan con el docente (83) y el seguimiento de las normas del centro durante la clase (85).

Los modelos se evalúan mediante chi-cuadrado robusto normalizado ($\chi^2 / d.f.$), para el que valores por debajo de 5 se consideran aceptables e inferiores a 3 muy buenos. También se utiliza el índice basado en los residuos *RMSEA* (Root Mean Square Error of Approximation), en el que los valores por debajo de .08 se consideran aceptables y por debajo de .06 óptimos. Finalmente se incluyen los índices *CFI* (Comparative Fit Index) y *TLI* (Tucker-Lewis Index), en los cuales un valor por encima de .9 se considera aceptable y por encima de .95 bueno. Como evidencia de validez, es suficiente un ajuste aceptable en la combinación los índices *RMSEA* y *CFI* (Hu y Bentler, 1999).

La dimensionalidad del constructo se refuerza con la interpretación de las correlaciones entre los factores latentes, así como con el estudio de la validez convergente, que se aborda a través de dos indicadores:

- 1) el promedio de la varianza explicada (*Average Variance Extracted, AVE*) o cociente entre la suma de las cargas factoriales estandarizadas al cuadrado y el número total de ítems de la dimensión. Valores por encima de .5 indican que la varianza explicada de cada factor es mayor que el error de medida o, en de otro modo, que se explica más del 50% de la variabilidad de las respuestas con el conjunto de ítems que componen el factor; y
- 2) y la Fiabilidad Compuesta (*FC*), calculada a partir de las cargas factoriales y de la varianza de error, también denominado índice Omega (McDonald, 1999). En este caso, valores superiores a .7 se interpretan como buena consistencia interna de los factores. Hair et al. (2010) aclaran que valores de *AVE* entre .4 y .5 son aceptables si los valores de *FC* son también aceptables.

Finalmente, se estudian los índices de modificación para identificar las posibles causas de los desajustes y construir el modelo final óptimo.

Además, se incluye la varianza explicada por el modelo de las respuestas a cada ítem (R^2).

Por otra parte, se prueba la invarianza del modelo final y sus parámetros entre los estudiantes de grado y los de máster, con un modelo factorial confirmatorio multigrupo siguiendo las recomendaciones de Brown (2015). En un primer paso se estima el modelo final de tres dimensiones de forma separada para cada grupo, grado y máster, empleando los mismos criterios de ajuste para la valoración de su calidad. A continuación, se ajustan una serie de modelos anida-

dos para probar la invarianza de la configuración (estructura de ítems y dimensiones igual), la invarianza de la métrica (pesos factoriales iguales) y la invarianza de la escala (intercepto y umbrales entre categorías equivalentes).

Para analizar la invarianza entre grupos se estudian las diferencias entre los valores de ajuste del modelo con más parámetros fijos comparado con el siguiente que tiene más parámetros libres. Debido a la sensibilidad del estadístico χ^2 al tamaño muestral y a las distribuciones no normales, que puede llevar a conclusiones de falta de invarianza cuando hay un empeoramiento significativo del ajuste, Cheung y Rensvold (2002) recomiendan como criterio para determinar falta de invarianza que se produzca un empeoramiento en el índice *CFI* de .01 o más. Además, Chen (2007) sugiere combinar los cambios en *CFI* con los que se producen en el índice *RMSEA*, donde un incremento igual o superior a .015 es evidencia también de falta de invarianza entre grupos. Todos los análisis del estudio confirmatorio se han realizado con el software MPlus 8 (Muthen & Muthen, 2017)

Para el análisis de la validez concurrente de los ítems, se calcula, con SPSS, CHAID (Chi Automatic Interaction Detection) (Kass, 1980), que aporta la significatividad de las diferencias de medias a través de χ^2 , y CART (Classification And Regression Trees) (Breiman et al., 1984), que permite jerarquizar los ítems según su importancia normalizada en un proceso supervisado. Son técnicas de clasificación que están demostrando ser útiles para la validación de medidas educativas (Álvarez y Asensio-Muñoz, 2020; Blanco et al., 2017). En este caso se utilizan como predictores los 17 ítems de la escala de Covarrubias (2014) y, como criterio, un ítem global de autoeficacia docente al que los estudiantes encuestados responden también en una escala Likert de 5 puntos.

Resultados

En primer lugar, se muestran en la Tabla 3 los estadísticos descriptivos univariados del conjunto de ítems, junto con las correlaciones bivariadas de cada reactivo con la variable criterio del segundo objetivo.

Tabla 3

Descripción de los ítems en la muestra de maestros y profesores en formación y correlaciones de Spearman con el ítem criterio.

	N	Media	DE	Rbo
1 comunicarte con los alumnos más difíciles? (70, 24, A1)	701	3.74	.784	.378**
2 ayudar a tus alumnos a pensar de manera crítica? (71, 22, A2)	740	4.05	.725	.276**
3 motivar a los alumnos que muestran poco interés por tu asignatura? (72, 19, A3)	741	4.11	.717	.420**
4 fomentar la creatividad de tus alumnos? (73, 23, A4)	741	4.11	.830	.302**
5 controlar el mal comportamiento en tu clase? (74, 9, C1)	741	3.74	.825	.359**
6 evaluar si tus alumnos comprenden lo que les explicas? (75, 7, B2)	741	4.01	.757	.273**
7 elaborar buenas preguntas para tus alumnos? (76, 3, B3)	741	3.93	.766	.300**
8 dar explicaciones o ejemplos adicionales cuando tus alumnos están confundidos? (77, 2, B4)	741	4.22	.739	.279**
9 responder a las preguntas difíciles que te hagan tus alumnos? (78, 5, B1)	741	3.79	.784	.351**
10 manejar a un alumno que presenta un mal comportamiento o que es "revoltoso"? (79, 11, C3)	739	3.69	.838	.365**
11 implementar estrategias de enseñanza alternativas en tu clase? (80, 4, D3)	738	4.01	.829	.333**
12 evitar que uno o unos pocos alumnos problemáticos te revienten la clase? (81, 13, C4)	741	3.57	.849	.321**
13 ajustar tus clases al nivel adecuado de cada uno de los alumnos? (82, 6, D1)	741	3.87	.805	.367**
14 manejar a esos alumnos que en un determinado momento llegaron a enfrentarse abiertamente a ti? (83, 14, C1)	739	3.63	.929	.334**
15 utilizar estrategias de evaluación distintas a las habituales? (84, 1, D1)	741	3.93	.874	.248**
16 que los alumnos sigan las normas del centro en tu clase? (85, 10, C2)	740	3.93	.735	.343**
17 ofrecer desafíos apropiados para los alumnos más capacitados? (86, 8, D4)	739	4.01	.811	.306**
18 ser el día de mañana un buen profesor? (87, criterio)	737	4.36	.754	1**

Nota. Junto al ítem entre paréntesis se indica el orden en el cuestionario completo, en la escala original y en la versión de Covarrubias (2014). Correlación significativa al 1% (**)

El ítem con una mayor media en percepción de autoeficacia, con una puntuación de 4.36 sobre 5, es el ítem global (87), seguido del 77, centrado en la capacidad de dar explicaciones o ejemplos adicionales cuando tus alumnos están confundidos. En el lado opuesto, aquellos ítems donde la muestra indica una menor autoeficacia media se refieren a la capacidad autopercebida para: evitar que uno o unos pocos alumnos problemáticos te revienten la clase (81), manejar a estudiantes que en un determinado momento llegaron a enfrentarse abiertamente al docente (83) y manejar a un estudiante que presenta un mal comportamiento (79).

Los ítems que más correlacionan con la percepción general de autoeficacia como futuro docente en la muestra total son la capacidad para motivar a los estudiantes que muestran

poco interés por la asignatura (72), con un valor de rho de .42, y la capacidad de comunicación con los estudiantes más difíciles (70), con .38.

Evidencias de validez estructural y convergente

No se identificaron casos perdidos, ni tampoco casos extremos mediante el estudio de las distancias de Mahalanobis. En cambio, el estadístico de curtosis multivariada de Mardia indica falta de normalidad del conjunto de ítems de la escala de actitud ($b2 = 402.330$; $N(b2) = 4.815$; $p = < .001$). Por ello, se utiliza la matriz de correlaciones policóricas y estimadores robustos basados en mínimos cuadrados (ULSMV) en los 3 modelos, cuyos índices de ajuste se ofrecen en la Tabla 4.

Tabla 4
Índices de ajuste de los modelos confirmatorios estimados.

Índice	Original	Covarrubias	Modelo propuesto
<i>Chi2</i>	662.742	596.597	476.418
<i>gl</i>	116	113	113
<i>p</i>	<.001	<.001	<.001
<i>Chi2/gl</i>	5.713	5.280	4.216
RMSEA	.083	.079	.068
RMSEA (LI 90%)	.077	.073	.062
RMSEA (LS 90%)	.089	.085	.075
RMSEA (<i>p</i>)	<.001	<.001	<.001
CFI	.924	.933	.950
TLI	.911	.920	.940

Considerando el ajuste global que proporciona el índice χ^2 normalizado, en ninguno de los casos se logra un valor óptimo (inferior a 3). El modelo propuesto es el único que logra un valor aceptable en este índice (inferior a 5).

El error de aproximación normalizado (RMSEA) obtiene valores de ajuste aceptables (por debajo de .08) en el modelo de Covarrubias (2014) y en el propuesto, aunque en el primero el límite superior está ligeramente por encima (.085). Los índices de ajuste comparativo CFI y TLI resultan por encima de .9 en los tres modelos testados, llegando a alcanzar un valor óptimo en el modelo propuesto (.95).

El estudio de la validez convergente en los tres modelos confirma que los factores están bien definidos, como se muestra en la Tabla 5.

Tabla 5
Índices de validez convergente (AVE y Fiabilidad).

Factores	Original	Covarrubias	Modelo
			Propuesto
AVE			
Estrategias EA	.387	.475	.417
Gestión del aula	.529	.529	.487
Implicación de estudiantes	.398	.398	.406
Eficacia en la atención a la singularidad		.400	
Fiabilidad Compuesta (FC)			
Estrategias EA	.833	.782	.809
Gestión del aula	.848	.848	.850
Implicación de estudiantes	.724	.724	.773
Eficacia en la atención a la singularidad		.726	

El índice AVE se sitúa por encima de .5 solo en la dimensión de gestión del aula en el modelo original y en el de

Tabla 7
Índices de ajuste del análisis factorial confirmatorio multigrupo.

	Grado	Master	Config.	Métrica	Escala	Métrica vs config.	Escalar vs config.	Escalar vs Métrica
<i>Chi2</i>	236.017	245.576	469.992	453.807	553.303			
<i>gl</i>	113	113	226	240	271			
<i>p</i>	<.001	<.001	<.001	<.001	<.001			
<i>Chi2/gl</i>	2.089	2.173	2.080	1.891	2.042			
RMSEA	.047	.075	.054	.049	.053			
RMSEA (LI)	.039	.063	.047	.042	.047			
RMSEA (LS)	.056	.088	.061	.056	.059			
RMSEA (<i>p</i>)	.677	<.001	.163	.576	.206			
CFI	.969	.908	.947	.953	.939			
Δ CFI						.006	-.009	-.014
Δ RMSEA						-.005	-.001	.004

Covarrubias (2014), aunque en estos dos modelos hay factores con valores de AVE ligeramente por debajo de .4. En el modelo propuesto, todos los factores logran un valor superior a .4, lo que, combinado con una FC cercana a .8, indica una validez convergente aceptable.

El factor mejor explicado en los modelos es gestión del aula, con un 50% aproximadamente de varianza explicada en el modelo propuesto y una FC de .85. En conjunto, los tres modelos logran explicar una varianza similar de las respuestas a cada uno de los ítems, como se muestra en la tabla 6, y en todos los casos es de .3 o superior. Los ítems mejor representados en los modelos son el 78 (responder a las preguntas difíciles que te hagan tus alumnos), 79 (manejar a un alumno que presenta un mal comportamiento o que es “revoltoso”) y 81 (evitar que uno o unos pocos alumnos problemáticos te revienten la clase), con un 60% aproximadamente de varianza explicada en los distintos modelos.

Tabla 6
Varianza explicada de cada uno de los ítems y varianza total explicada.

	R ²		
	Original	Covarrubias	Modelo propuesto
70, 24, A1	.476	.476	.437
71, 22, A2	.359	.358	.39
72, 19, A3	.462	.463	.495
73, 23, A4	.295	.0295	.319
74, 9, C1	.471	.47	.47
75, 7, B2	.342	.4	.336
76, 3, B3	.343	.398	.333
77, 2, B4	.418	.479	.435
78, 5, B1	.525	.623	.59
79, 11, C3	.604	.604	.589
80, 4, D3	.415	.455	.478
81, 13, C4	.624	.624	.547
82, 6, D1	.437	.479	.494
83, 14, C1	.539	.539	.472
84, 1, D1	.31	.34	.35
85, 10, C2	.409	.409	.407
86, 8, D4	.303	.326	.316
Total	.31	.455	.439

El análisis multigrupo muestra la equivalencia del modelo propuesto tanto en los estudiantes de grado como en los de máster (tabla 7).

Los AFC por separado para los grupos (grado vs máster) lograron valores de ajuste aceptables ($RMSEA < .08$ y $CFI > .9$), aunque el ajuste es mejor con la muestra de estudiantes de grado ($RMSEA < .06$ y $CFI > .95$).

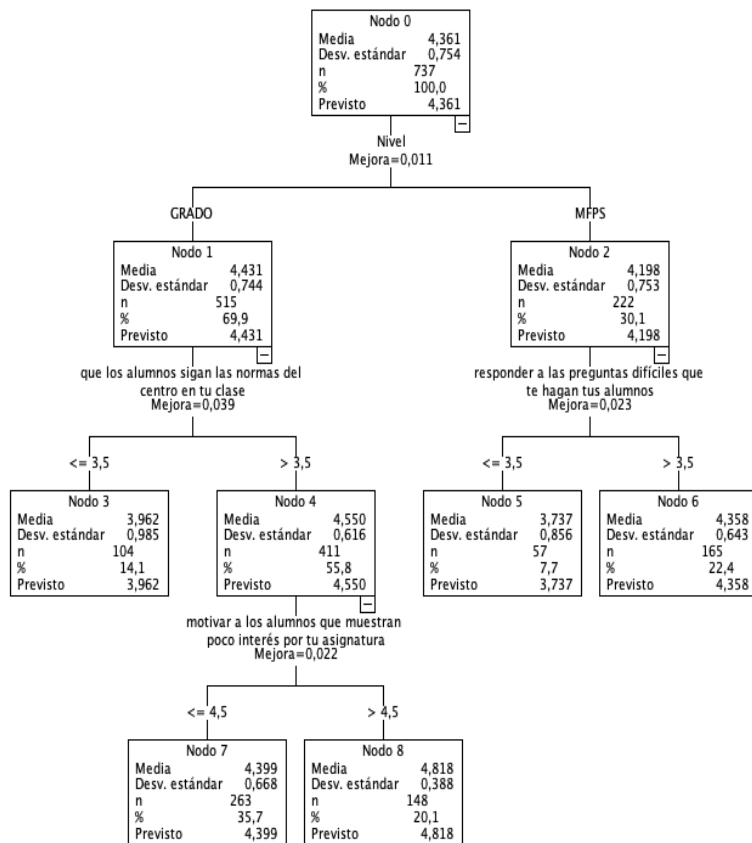
El modelo de configuración multigrupo que no lleva a cabo ninguna restricción en los parámetros de los grupos, excepto la estructura de ítems y factores, mostró buenos índices de ajuste $RMSEA$ y CFI . El modelo que fija las cargas factoriales para mantenerlas iguales entre los grupos (métrica) mejora ligeramente el ajuste del modelo de configuración, aunque los cambios son mínimos (una mejora de aproximadamente .005 en cada índice). Por tanto, la invarianza de la métrica queda demostrada.

En el modelo escalar, donde los interceptos y los umbrales de los ítems se mantienen iguales entre grupos, también se logra un buen ajuste. Sin embargo, los cambios en el ajuste respecto al modelo que prueba la invarianza métrica, empeoran ligeramente, sobre todo CFI que supera el umbral recomendado de .01. El cambio en $RMSEA$ es inferior a .005. Por consiguiente, la invarianza escalar no puede asegurarse.

Evidencias de validez concurrente

Tomando como criterio el ítem de alta inferencia (87 en el cuestionario completo), con CHAID emergen como ítems que discriminan significativamente entre alguno de los nodos para la muestra completa: el 72, el 81, el 82, el 85 y el 77 del cuestionario completo. El algoritmo CART coincide en que los ítems 72 y 85 son los más importantes cuando se trata de explicar el SAD global en la muestra total. En la figura 1, incluyendo como primera variable forzosa el nivel (grado o máster), se observa que los estudiantes de grado con menor autoeficacia docente percibida (media = 3.962), se sitúan en el nodo 3: son estudiantes que se puntúan con 3 o menos en el ítem 85, relativo a la capacidad para conseguir que los estudiantes sigan las normas. Los estudiantes de grado con mayor autoeficacia global autopercibida (media = 4.818) se encuentran en el nodo 8: se otorgan 4 o más en el ítem 85 y se puntúan con 5 en el ítem 72, esto es, se sienten altamente capaces de motivar a los estudiantes que muestran poco interés. Para el grupo de los estudiantes de postgrado, los que se sienten más eficaces (media = 4.358) están en el nodo 6 y son los que se sienten capaces de responder a las preguntas difíciles que se les pueda plantear en clase, esto es, se han puntuado con un 4 o más en el ítem 78.

Figura 1
Diagrama de árbol CART para la muestra total forzando el nivel como primera variable.
ser el día de mañana un buen profesor



En la submuestra de estudiantes de grado la importancia relativa de las variables en la explicación del criterio según el algoritmo CART se ofrece en la figura 2. En ella, tras los ítems 72 y 85 que aparecen también en el diagrama de árbol que se ofrece en la figura 1, se sitúan los ítems 80, 82 y 86,

que emergen como significativos en el árbol CHAID que se ofrece en la figura 3 porque discriminan entre nodos, tomando como valores de corte el 3 o el 4 de la escala Likert según el caso.

Figura 2

Gráfica CART de la importancia de las variables en la submuestra de estudiantes de grado.

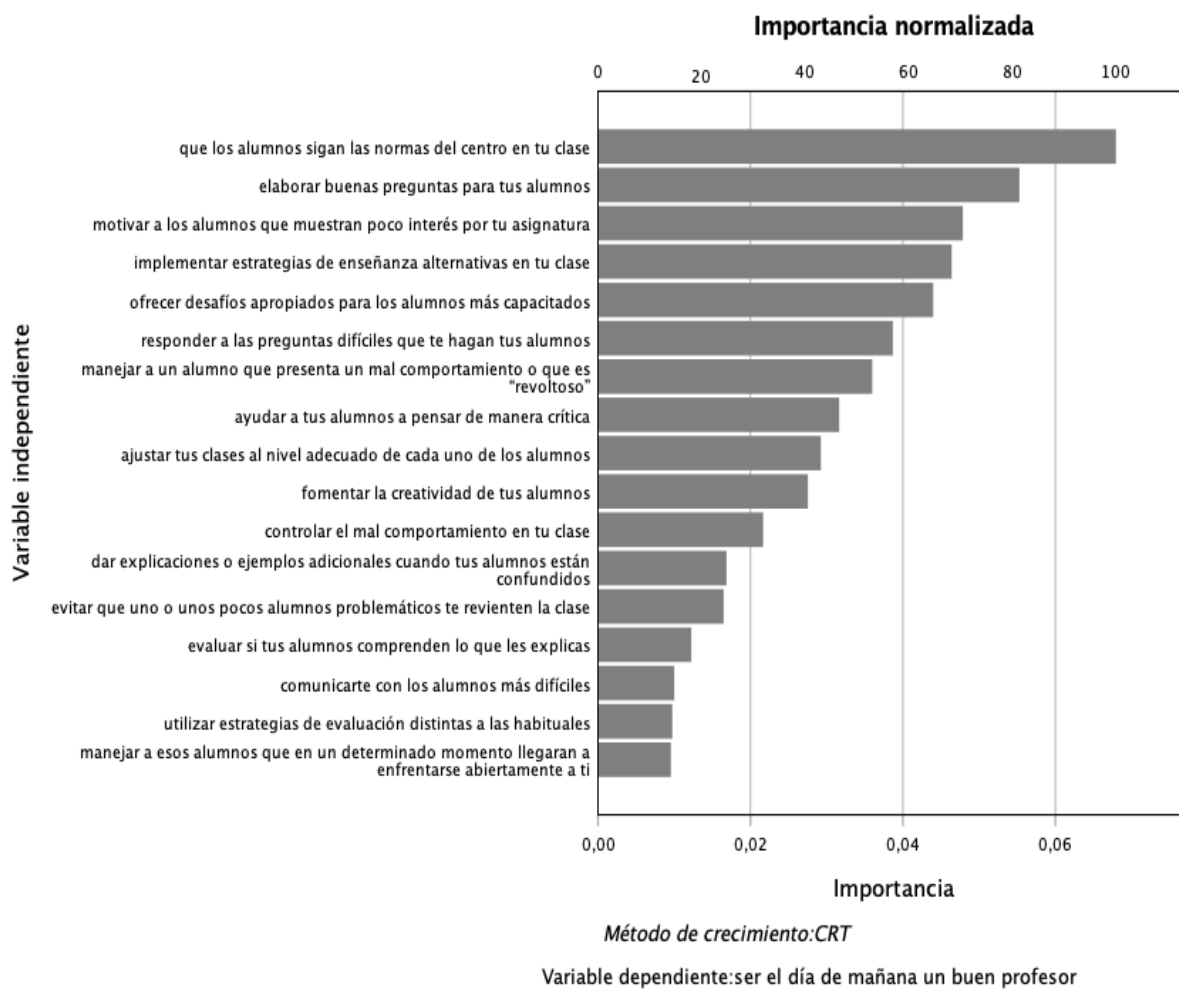
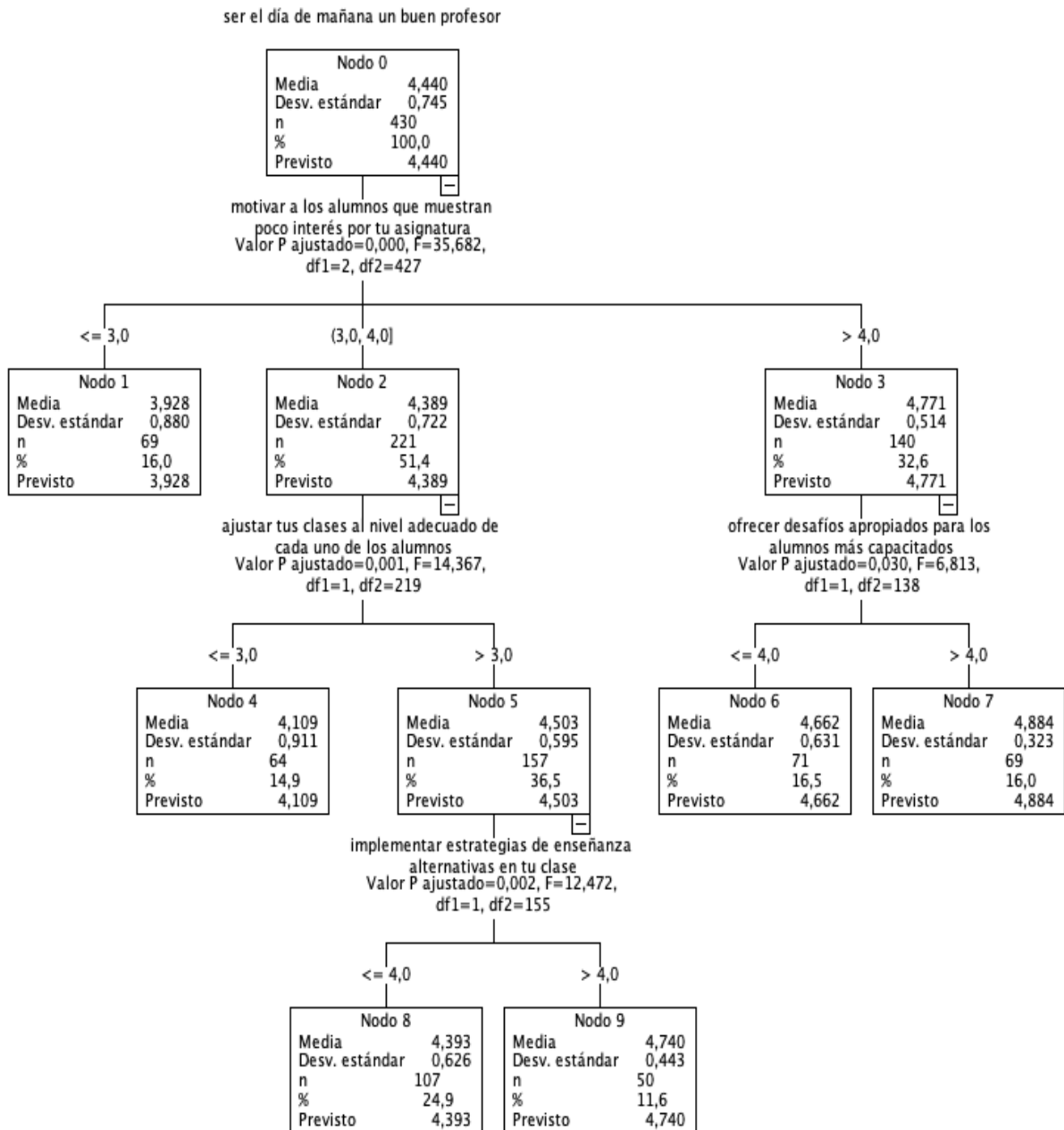


Figura 3
Diagrama de árbol CHAID en la submuestra de estudiantes de grado.

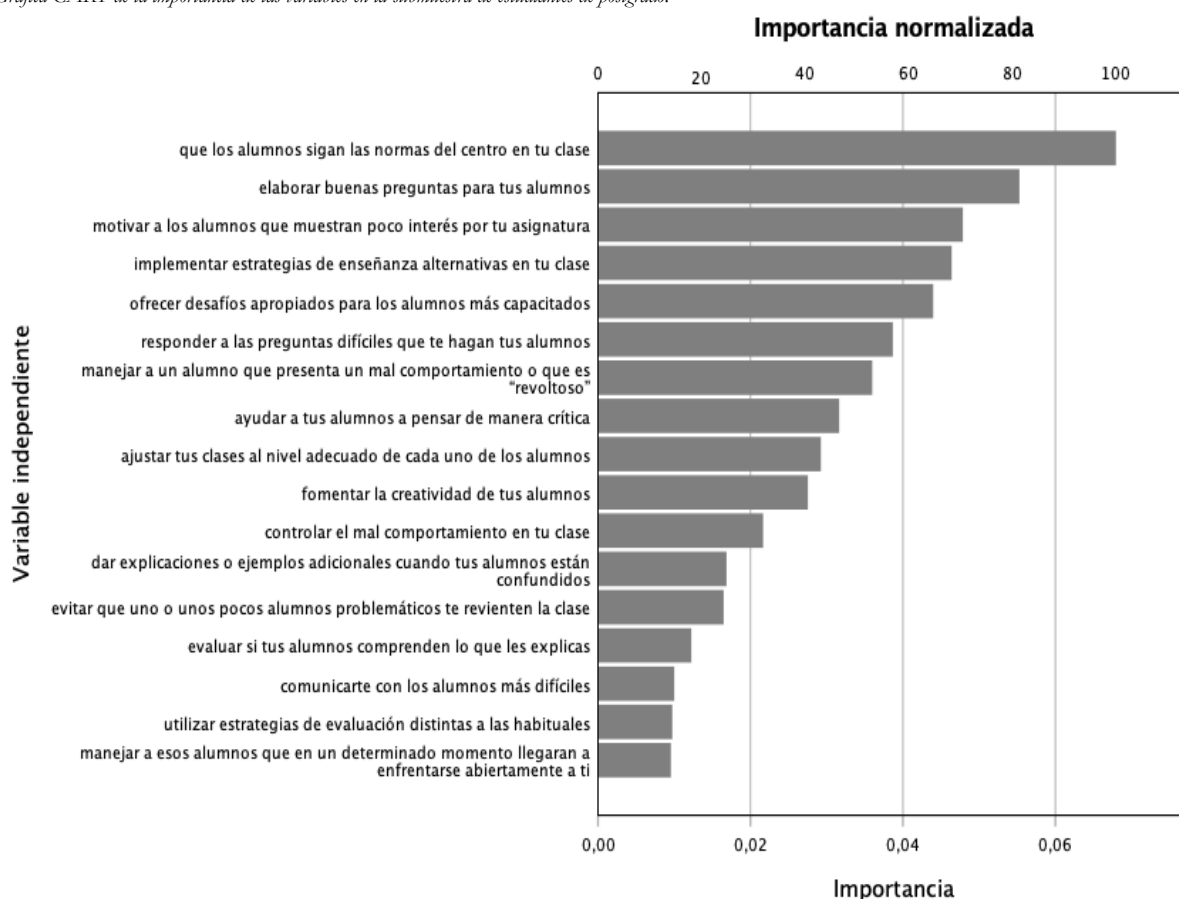


En los estudiantes de máster el orden de importancia de las variables cambia, como se puede observar en la figura 4, de modo que la variable claramente más importante en su re-

lación con el criterio de validez empleado es el ítem 78 que se refiere a su capacidad para responder preguntas difíciles que puedan hacer los estudiantes.

Figura 4

Gráfica CART de la importancia de las variables en la submuestra de estudiantes de postgrado.



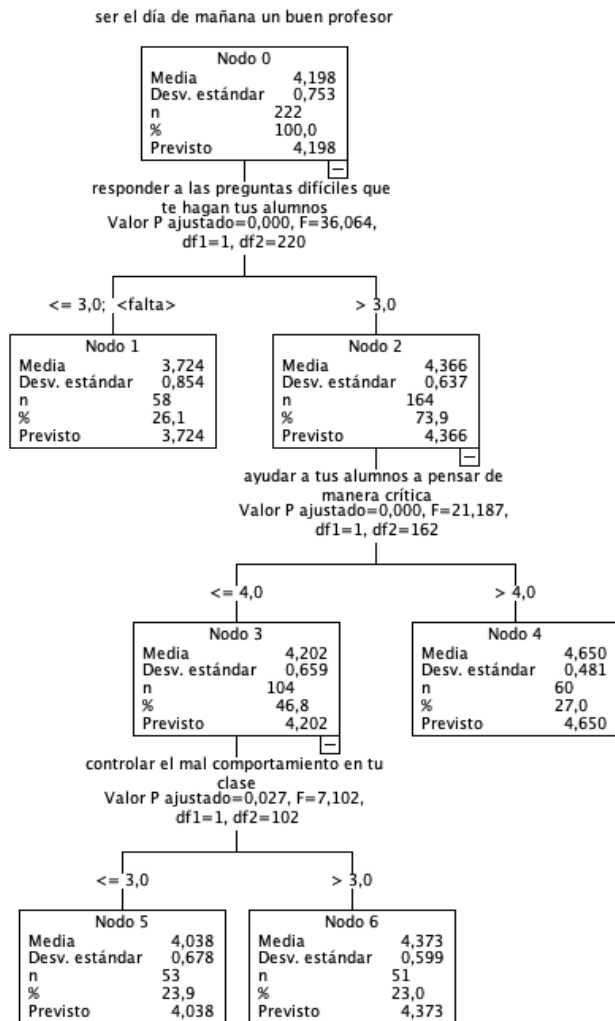
Método de crecimiento: CRT

Variable dependiente: ser el día de mañana un buen profesor

El árbol para los estudiantes de máster encuestados se ofrece en la figura 5 en el que emerge de manera robusta que el ítem 78 es el ítem más importante en la definición del concepto de autoeficacia global en este subgrupo. En este diagrama de árbol se aprecia que un punto de corte de 4 en

el ítem 71 (capacidad para ayudar a los estudiantes a pensar de manera crítica) permite dividir el nodo 2 e identificar a un subgrupo de 60 estudiantes que conforman el nodo 4, en el que se hallan los estudiantes de máster con un nivel de autoeficacia global percibida más alta (media = 4.65).

Figura 5
Diagrama de árbol CHAID en la submuestra de estudiantes de máster.



Discusión

Los resultados de la validación a los que se llega en este estudio avalan un constructo estructurado en tres factores, lo que coincide con la propuesta original y la mayor parte de las investigaciones que han utilizado este modelo de medida. Sin embargo, hay ítems que no responden con claridad ni a la estructura original ni a la propuesta por Covarrubias (2014). Por una parte, la autoeficacia percibida en la utilización de estrategias de enseñanza y evaluación, que en el cuestionario original forman parte del manejo de las estrategias de EA (ítems 80 y 84), en la muestra de profesores en activo chilenos son manifestación observada de una variable latente nueva. En la muestra de estudiantes españoles el modelo mejora cuando se consideran variables observadas ligadas al factor de implicación de los estudiantes, lo que puede justificarse porque utilizar estrategias novedosas normalmente va asociado al fomento de la motivación. Las estrategias adapta-

tivas (ítems 82 y 86), que Covarrubias saca del factor original para crear un cuarto factor de atención a la diversidad, en el modelo propuesto en esta investigación se asocian satisfactoriamente con la dimensión original. Por último, el ítem 70 (conseguir llegar a los estudiantes más difíciles), tanto en el modelo original como en el de Covarrubias (2014) forma parte del factor de implicación de los estudiantes, mientras que en el modelo al que se llega en el presente estudio forma parte de la dimensión de gestión del aula, lo que tiene sentido teórico y puede explicarse por la traducción del verbo principal, que en inglés es polisémico.

En este punto, cabe señalar que, a lo escurridizo o difuso del constructo hay que unir el problema que supone en los estudios de validación la traducción a un idioma distinto del original. Covarrubias (2014), Covarrubias y Mendoza (2016) y Burgueño et al. (2019) utilizan el procedimiento de traducción inversa para llegar a la versión del TSES en español, mientras que el resto de los investigadores que han utilizado este instrumento con hispanohablantes no especifican el procedimiento seguido para la adaptación.

Por otra parte, la medida del constructo en español se puede ver afectada por variables de contexto como el país (Chile, Colombia, España, República Dominicana o Perú) o el tipo de docentes. En cuanto al país de la muestra, además de las diferencias propias de cada nacionalidad, hay que considerar que en cada uno el uso del español como lengua es distinto. En cuanto al tipo de docentes, los trabajos de validación realizados hasta el momento con muestras hispanoparlantes se han realizado fundamentalmente con profesores en activo. Solo Burgueño et al. (2019) trabajan con profesores en formación de un único modelo formativo. En este estudio se trabaja con futuros docentes de una universidad española procedentes de dos modelos de formación del profesorado: 1) los estudiantes que llegarán a ser maestros de Educación Infantil o Primaria (de 0 a 12 años), que se forman en sendos grados, y 2) los estudiantes que en un año estarán capacitados para dar enseñanzas en Educación Secundaria, Bachillerato y Formación Profesional (MFPS), que se forman en un máster al que se accede desde diversos grados universitarios en los que previamente se han preparado en la parte disciplinar. En este sentido, el trabajo aquí presentado viene a cubrir colectivos no estudiados previamente.

En lo referente a validez criterial, todos los ítems incluidos en el análisis son relevantes en la predicción del ítem global de alta inferencia incluido, la autoeficacia general, entendida como la autopercepción de la capacidad de llegar a ser un buen profesor, dado que todas las correlaciones son significativas, como se muestra en la tabla 3. Los árboles de decisión matizan esta información y nos descubren que los ítems más ligados con el tipo de percepción de la autoeficacia docente que tienen los estudiantes de grado, futuros maestros de EP y EI, son la capacidad para conseguir que los chicos sigan las normas de clase (85) y la capacidad que sienten que tienen de motivar (72) a los estudiantes que sienten poco interés por el trabajo escolar. En cambio, en el nivel de postgrado, el ítem más relevante es el 78 en el que se

plantea la capacidad que sienten que tienen de responder a preguntas difíciles.

Conclusiones

Con el estudio realizado se logran los objetivos fijados, ya que se obtienen unos índices de ajuste aceptables para la dimensionalidad de la autoeficacia docente propuesta y se llega a un modelo consistente y válido en los dos subgrupos considerados, estudiantes de grado y de máster. Por otra parte, se ofrece evidencia de la utilidad de los ítems para predecir la percepción que los estudiantes tienen de su autoeficacia docente global. Una aportación de interés es la importancia diferencial que determinados ítems tienen en la explicación del criterio elegido según los estudiantes sean de grado o de máster. En este sentido, destaca el valor que tiene el conocimiento de la materia que se enseña en la autoeficacia de los estudiantes de máster, en su mayoría futuros profesores de secundaria, mientras que, en los estudiantes de grado, futuros maestros de niños entre 0 y 12 años, lo que más pesa en su SAD son las normas y la motivación. Aunque el constructo responde a la misma estructura factorial, como se ha mostrado en el AFC multigrupo, los estudiantes de grado y de máster tienen perfiles diferenciales de autoeficacia y los ítems que componen el cuestionario contribuyen a discriminarlos.

Referencias

- AERA, APA, & NCME (2014). *Standards for Educational and Psychological Testing*. APA.
- Álvarez, M. M., & Asensio-Muñoz, I. I. (2020). Evidence of validity of a generic competencies measure. *Educación XX1*, 23(2), 337-366. <https://doi.org/10.5944/educxx1.25896>
- Asensio-Muñoz, I., & Ruiz de Miguel, C. (2017). Measurement and assessment of beliefs about the profession at initial teachers training. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 20(3), 79-91. <https://doi.org/10.6018/reifop.20.3.265231>
- Azzi, R. G., & Polydoro, S. A. (Eds.). (2006). *Auto-eficácia em diferentes contextos [Self-efficacy in different contexts]*. Alínea.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215. <https://doi.org/1.1037/0033-295X.84.2.191>
- Bermejo, L., & Prieto, M. (2005). Educational uneasiness and teacher self-efficacy beliefs. *Revista Española de Pedagogía*, 42(232), 493-51. <https://bit.ly/3uaUU68>
- Blanco, Á., Asensio, I., Carpintero, E., Ruiz de Miguel, C., & Expósito, E. (2017). Applications of hierarchical segmentation in measurement and evaluation of educational programs. Examples with a financial education program. *Educación XX1*, 20(2), 235-257. <https://doi.org/10.5944/educXX1.14464>
- Breiman, L., Friedman, J. H., Olshen R. A., & Stone, C. J. (1984). *Classification and regression trees*. Chapman & Hall/CRC.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2th ed.). The Guilford Press.
- Bueno-Álvarez, J. A., Martín-Martín, M. y Asensio-Muñoz, I. (2022). Self-Efficacy in Future Teachers: Systematic Review in Iberoamerica 2015-2020. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 20(4), 31-50. <https://doi.org/10.15366/reice2022.20.4.002>
- Burgueño, R., Sicilia, A., Medina-Casabón, J., Alcaraz-Ibañez, M., & Lirola, M. J. (2019). Psychometry of the teacher's sense of efficacy scale in Spanish teachers' education. *The Journal of Experimental Education*, 87(1), 89-10. <https://doi.org/1.1080/00220973.2018.1423542>
- Cain, M. K., Zhang, Z., & Yuan, K. H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior research methods*, 49, 1716-1735. <https://doi.org/10.3758/s13428-016-0814-1>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464-504. <https://doi.org/1.1080/10705510701301834>
- Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Education Research and Review*, 15, 1-16. <https://doi.org/1.1016/j.edurev.2015.02.001>
- Cheung, G. W., & Rensvold, R. B. (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255. https://doi.org/1.1207/S15328007SEM0902_5
- Covarrubias, C. G. (2014). *The sense of self-efficacy in a sample of Chilean teachers* (Tesis Doctoral). Universidad Complutense de Madrid. <https://bit.ly/3feiElr>
- Covarrubias, C. G., & Mendoza, M. C. (2016). Adaptation and validation of the Self efficacy Sense Questionnaire in a sample of Chilean teachers. *Universitas Psychologica*, 15(2), 97-108. <https://doi.org/1.11144/Javeriana.upsy15-2.avcs>
- Dominguez-Lara, S., Fernández-Arata, M., Merino-Soto, C., Navarro-Loli, J., & Calderón-De la Cruz, G. (2019). Teacher's Self-Efficacy Scale: structural analysis and measurement invariance in Peruvian teachers of public schools. *Revista Argentina de Ciencias del Comportamiento*, 11(3), 61-72. <https://doi.org/1.32348/1852.4206.v11.n3.24624>
- Friedman, I. A., & Kass, E. (2002). Teacher self-efficacy: a classroom-organization conceptualization. *Teaching and Teacher Education*, 18(6), 675-686. [https://doi.org/1.1016/S0742-051X\(02\)00027-6](https://doi.org/1.1016/S0742-051X(02)00027-6)
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Prentice Hall.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. <https://doi.org/1.1080/10705519909540118>

- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 119-127. <https://doi.org/1.2307/2986296>
- Kleinsasser, R. C. (2014). Teacher efficacy in *Teaching and Teacher Education*. *Teaching and Teacher Education*, 44, 168-179. <https://doi.org/1.1016/j.tate.2014.07.007>
- Lera, M. J., León-Pérez, J. M., & Ruiz-Zorrilla, P. (2021). Adaptation of the Teacher Efficacy Scale to measure effective teachers' educational practices through students' ratings: A multilevel approach. *Psicothema* 33(3), 509-517. <https://doi.org/1.7334/psicothema202.262>
- Li, C. H. (2014). *The performance of MLR, USLMV, and WLSMV estimation in structural regression models with ordinal variables*. Michigan State University. <https://bit.ly/3FTU60r>
- Martínez-Luque, C. M., Hervás-Gómez, C., & Román-Graván, P. (2017). Experience in the university field: self-efficacy and teacher motivation. *International Journal of Educational Research and Innovation*, 8, 175-184. <https://bit.ly/3uaX7ys>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
- Muthen, L. K., & Muthen, B. O. (2017). MPLUS (Version 8). [Software].
- Pajares, F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332. <http://doi.org/1.3102/00346543062003307>
- Perandones, T. M., & Castejón, J. L. (2007). Estudio correlacional entre personalidad, inteligencia emocional y autoeficacia en profesorado de educación secundaria y bachillerato [Correlational study between personality, emotional intelligence and self-efficacy in Secondary and High School teachers]. In A. Jiménez & M. A. Lou (Eds.), *V Congreso Internacional "Educación y Sociedad". La Educación: Retos del S. XXI* (pp. 1-8). Ilustre Colegio de Doctores y Licenciados en Ciencias y en Letras de Granada, Almería y Jaén. <https://bit.ly/2RLkGk6>
- Perandones, T. M., & Herrera, L. (2017). Teacher self-efficacy and personal strengths and virtues in teachers from Dominican Republic. *International Journal of Developmental and Educational Psychology*, 1(2), 387-396. <https://doi.org/1.17060/ijodaep.2017.n1.v3.1006>
- Rodríguez, S., Nuñez, J. C., Valle, A., Blas, R., & Rosario, P. (2009). Autoeficacia Docente, motivación del profesor y estrategias de enseñanza [Teacher self-efficacy, teacher motivation, and teaching strategies]. *Escritos de Psicología*, 3(1), 1-7. <https://bit.ly/3wXbEmj>
- Salanova, M., Grau, R., Martínez, I. M., Cifre, E., Llorens, S., & García-Renedo, M. (Eds.), (2004). *Nuevos horizontes en la investigación sobre autoeficacia [New horizons in self-efficacy research]*. Servicio de Publicaciones de la Universidad Jaume I.
- Salas-Rodríguez F., Lara, S., & Martínez, M. (2021). Spanish version of the Teachers' Sense of Efficacy Scale: An adaptation and validation study. *Frontiers in Psychology*, 12, 714145. <https://doi.org/1.3389/fpsyg.2021.714145>
- Suárez-Escobar, S. (2018). Relación entre la autoeficacia docente y la autoestima en profesores que laboran en un grupo de colegios colombianos [Relationship between teacher self-efficacy and self-esteem in teachers working in a group of Colombian schools]. *Revista Internacional de Estudios en Educación*, 18(1), 46-6. <https://doi.org/1.37354/riee.2018.179>
- Tschannen-Moran, M., & McMaster, P. (2009). Sources of self-efficacy: four professional development formats and their relationship to self-efficacy and implementations of a new teaching strategy. *Elementary School Journal*, 110(2), 228-245. <https://doi.org/1.1086/605771>
- Tschannen-Moran, M., Salloum, S.J., & Goddard, R. D. (2015). Context matters. The influence of collective beliefs and shared norms. In H. Fives, & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 301-316). Routledge. <https://doi.org/1.4324/9780203108437>
- Tschannen-Moran, M., Woolfolk-Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: its meaning and measure. *Review of Educational Research*, 68(2), 202-248. <https://doi.org/1.3102/00346543068002202>
- Tschannen-Moran, M., & Woolfolk-Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783-805. [https://doi.org/1.1016/S0742-051X\(01\)00036-1](https://doi.org/1.1016/S0742-051X(01)00036-1)
- Uriarte, J. D., Pegalajar, M., León-López, J. M. de, & Galindo, H. (2019). Relationships between attitudes towards inclusive education, self-efficacy and teaching resilience. *International Journal of Developmental and Educational Psychology*, 2(1), 75-86. <https://doi.org/1.17060/ijodaep.2019.n1.v2.1452>
- Xia, Y. (2016). *Investigating the chi-square-based model-fit indexes for WLSMV and ULSMV estimators*. <https://bit.ly/3FJRldZ>
- Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, 86(4), 981-1015. <https://doi.org/1.3102/0034654315626801>