anales de psicología / annals of psychology
2023, vol. 39, nº 2 (may), 252-264
https://doi.org/10.6018/analesps.522321

ISSN print: 0212-9728. ISSN online: 1695-2294.
https://revistas.um.es/analesps

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

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

Universidad Complutense de Madrid (Spain)

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

Abstract: Teachers' sense of self-efficacy is a hard construct to measure but is important in view of to 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.

Introduction

The sense of teacher self-efficacy (TSE) is one of the qualities that most and best predicts teachers' educational performance, as it is an important indicator of an individual teacher's level of professional satisfaction and is a good predictor of their students' potential attainment (Chesnut & Burley, 2015; Zee & Koomen, 2016). As such, there is interest in it as a research topic in educational psychology.

As a construct, it is part of the framework of social learning of Bandura (1977), which makes it possible to explain human beings' behaviour in multiple contexts (Azzi & Polydoro, 2006; Salanova et al., 2004).

TSE is defined as individual teachers' beliefs about their capacity to tackle the everyday tasks that the teaching–learning process involves, at an individual and group level. As a belief, it is a complex construct with cognitive and affective components, which shapes how individuals interpret their professional position and engage with it (Asensio-Muñoz & Ruiz de Miguel, 2017), and so it acts as a mediator between teaching knowledge and action (Pajares, 1992). The complexity mentioned above increases as this sense is also affected by the context in which the teaching activity is performed (Friedman & Kass, 2002).

Tschannen-Moran et al. (1998; Tschannen-Moran & Woolfolk Hoy, 2001) present the theoretical model of TSE that is currently most widely accepted, thanks to its explanatory and predictive value. In it, TSE is characterised by two components that mutually influence one another and, at a behavioural and a cognitive level, shape the everyday work of the teacher: analysis of the task and self-evaluation of personal competence (Tschannen-Moran & McMaster, 2009).

TSE is affected not only by in-class teaching tasks, but also by the academic and organisational context of the centre: the teacher simultaneously acts as leader of the classroom at the task level and as an employee of the centre at the relational level (Tschannen-Moran et al., 2015).

To evaluate this proposal, Tschannen-Moran and Woolfolk Hoy (2001) developed the Ohio State Teacher Efficacy Scale (OSTES), which, after various analyses, gave rise to the Teachers' Sense of Efficacy Scale (TSES). This comprises 24 items (with a reduced version of 12), grouped into three factors: efficacy for instructional strategies (8 items); efficacy for classroom management (8 items); and efficacy for student engagement (8 items). This is currently the most widely used instrument for evaluating this construct internationally (Kleinsasser, 2014) and in research from Latin America (Bueno-Álvarez et al., 2022).

In the Spanish-speaking context, on which we focus in this article, applications of the TSES are reported by Bermejo and Prieto (2005), Perandones and Castejón (2007) and Rodríguez et al. (2009), but the first version to have been validated is that of Covarrubias (2014) and Covarrubias and Mendoza (2016). These authors reduce the number of items to 17 and identify another factor, proposing an instrument with a four-part factorial structure. Other research carried out after Covarrubias (2014) either does not perform a psychometric study (Martínez-Luque et al., 2017; Perandones &

*** 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 (Spain).
E-mail: alvarez@ucm.es
(Article received: 26-04-2022; revised: 16-09-2022; accepted: 10-10-2022).

Herrera, 2017; Suárez-Escobar, 2018; Uriarte et al., 2019) or reproduces the same original factorial structure (Burgueño et al., 2019; Domínguez-Lara et al., 2019; Lera et al., 2021; Salas-Rodríguez et al., 2021). Table 1 provides a summary description of these studies.

Table 1
Applications of TSES in Spanish.

Authors	Sample	Form	Structure	Validation
Bermejo and Prieto (2005)	71 Spanish compulsory secondary education teachers	TSES in Spanish, no other information, 9-point answer scale	3 subscales, 12 items	
Burgueño et al. (2019)	358 Spanish master's level secondary teacher training students	TSES-24, TSES-12 and TSES-11. Back translation with 9-point answer scale	3 subscales, 24, 12 and 11 items respectively	Validity: Exploratory Factor Analysis (EFA); Confirmatory Factor Analysis (CFA): max. likelihood, bootstrapping. Goodness of fit: χ2, ratio between chi squared and degrees of freedom (*χ2/df*), *CFI, IFI, TLI, RMSEA, SRMR*
Covarrubias (2014); Covarrubias and Mendoza (2016)	544 inservice Chilean teachers	TSES back translation with 5-point answer scale	4 subscales, 17 items	Construct: EFA: CFA: max. likelihood, bootstrapping. Goodness of fit: *χ2*, ratio between chi squared and degrees of freedom (*χ2/df*), *CFI, IFI, RMSEA, SRMR*. Reliability: Cronbach's alpha
Domínguez-Lara et al. (2019)	347 Peruvian non-higher education teachers	TSES in Spanish (TSE) with 5-point answer scale	3 subscales, 24 items	Validity: Structural equations (WLSMV) *CFI, RMSEA, ESEM:IFS* Reliability: Cronbach's alpha
Lera et al. (2021)	2242 children from two autonomous regions of Spain	Translation (no report of how it was done) and adaptation of the OSTES so that children can answer it on a 5-point scale	3 subscales, 24 items	EFA Ordinal CFA MCFA Reliability: omega
Martínez-Luque et al. (2017)	52 Spanish university teachers	TSES in Spanish, no report of how the translation was done, 9-point answer scale	3 subscales, 24 items	Not calculated, but the alpha of another study is reported
Perandones and Castejón (2007); Perandones and Herrera (2017)	564 Dominican EY, PT teachers	TSES-short form. Direct translation by Perandones and Castejón (2017) with 9-point answer scale	3 subscales, 24 items	Construct validity: CFA principal components and varimax, KMO and Bartlett Reliability: Cronbach's alpha
Rodríguez et al. (2009)	95 Spanish university teachers	TSES in Spanish (nothing else specified)	3 subscales	
Salas-Rodríguez et al. (2021)	190 Mexican PT and secondary teachers	TSES-12. Back translation with 9-point answer scale	3 subscales	Construct: CFA: max. likelihood. Goodness of fit: *χ2*, ratio between chi squared and degrees of freedom (*χ2/df*), *CFI, TLI, RMSEA, SRMR*. Reliability: Cronbach's alpha
Suárez-Escobar (2018)	86 Colombian non-higher education teachers	TSES in Spanish, no information about how the translation was done, 9-point answer scale	3 subscales, 24 items	Reliability: Cronbach's alpha
Uriarte et al. (2019)	300 teachers from the Dominican Republic	TSES in Spanish, no information about how the translation was done, 9-point answer scale	3 subscales, 24 items	Reliability: Cronbach's alpha

Given the use made of the TSES in Spanish, the variety of translations of it and the fact that teacher self-efficacy is possibly still an elusive construct (Tschannen-Moran & Woolfolk Hoy, 2001), it would be of value to continue to gather evidence about the structural validity that Covarrubias and Mendoza (2016) attain with Chilean teachers. In the present study, we work with a sample of trainee teachers from Spain and use ordinal confirmatory factor analysis (CFA), which is usually more appropriate for data from Likert-type scales. Likewise, it is intended to progress in the study of the concurrent validity of the scale by using decision trees, which are also suitable for ordinal data (Blanco et al., 2017). Consequently, a validation study is presented below, which is a process of accumulation of evidence about a measurement

instrument for the purposes it pursues, in which the characteristics of the data and the objectives of the instrument determine the most appropriate type of methodological approach in each case (AERA, APA & NCME, 2014). The aim of this work is to consider in depth the operative definition of the construct in Spanish and the validity of the measurement instrument proposed by Covarrubias and Mendoza (2016), gathering new evidence to support 1) the dimensionality of teacher self-efficacy, and the convergence and internal consistency of the proposed measurement model and the possible differences between undergraduate and master's students; and 2) the utility of the items from the adapted TSES for identifying the items that best predict students' perception of their overall teacher self-efficacy, taking into account whether they are undergraduate or master's students.

Method

Participants

The data are taken from a convenience sample of 744 students at the Universidad Complutense de Madrid who are in the first year of undergraduate degrees in Primary School Teaching (PT) (28%) and Early Years Education (EY) (19.7%), in the double degrees in PT with EY or with Pedagogy (18.2%) and the master's in Secondary Teacher Training (MSTT) (34%) with various specialisms. The mean age was 21.13 years (SD = 4.573). Women represent 73.8% of the sample.

Instrument

Covarrubias' version (2014) of the Teachers' Sense of Efficacy Scale (TSES) was used. This is an adaptation of the Teachers Self Efficacy Scale of Tschannen-Moran and Woolfolk-Hoy (2001). Covarrubias (2014) does a back translation of the original instrument. A five-point scale is used in the adaptation to Spanish and it is applied to a sample of inservice teachers from Chile. The exploratory and confirmatory factor analyses of the adapted instrument, published by Covarrubias and Mendoza (2016) support the adequacy of a measurement with four factors and 17 items in the Spanish-speaking sample. Covarrubias (2014) and Covarrubias and Mendoza (2016) describe the structure that emerges in the Chilean sample, which is as follows:

A. Efficacy in student engagement. Factor related to teachers' capacity to motivate students and get them involved in teaching and learning (A1, A2, A3, A4, which, as Table 3 shows, are equivalent to items 70, 71, 72 and 73 in our study and items 24, 22, 19 and 23 from the original scale)
B. Efficacy in instructional strategies. Factor focussed teachers' capacity to use different strategies and/or methodologies to teach and inspire learning in the students. (B1, B2, B3, B4, which, as Table 3 shows, are equivalent to items 78, 75, 76 and 77 in our study and items 5, 7, 3 and 2 from the original scale)
C. Efficacy in classroom management. Factor linked to teachers' capacity to handle or manage behaviour, discipline and order, favouring a positive atmosphere in class. (C1, C2, C3, C4 and C5, which, as Table 3 shows, are equivalent to items 74, 85, 79, 81 and 83 in our study and items 9, 10, 11, 13 and 14 from the original scale)
D. Efficacy in the attention to students' singularity. Factor connected to teachers' capacity to adapt the teaching to the specific learning needs and/or demands of each student. (D1, D2, D3, D4, which, as Table 3 shows, are equivalent to items 82, 84, 80 and 86 in our study and items 6, 1, 4 and 8 from the original scale)

This Spanish version, reduced to 17 items, is the one used in the present study.

Procedure

The instrument was administered in person by teachers of groups who agreed to collaborate in the investigation. Once general instructions had been given, the participants gave informed consent and voluntarily completed the teacher self-efficacy questionnaire, as well as answering other sociodemographic and high-inference questions that are relevant for the study.

Statistical and psychometric analysis

To collect evidence on the structure of the construct and meet the first objective, we checked for missing values and multivariate nonnormality through Mardia's multivariate measure of kurtosis, using the *b2* measure (Cain et al., 2017) and identified outliers by estimating Mahalanobis distances, then calculated the values of the parameters of the confirmatory models based on a matrix of polychoric correlations and the method of estimating robust unweighted least squares (ULSMV) (Li, 2014; Xia, 2016).

Owing to the ordinal nature of the items, the link function is a probit model that estimates intercepts and thresholds. These thresholds identify the level or score necessary in the dimension to put it in the different response categories. In other words, the first step displays the level to indicate category 1 with regards to 0, the second step category 2 with regards to 1, and so on successively.

Three measurement models were tested. The first replicates the original three-dimensional structure of Tschannen-Moran and Hoy (2001), the second is the four-factor model of Covarrubias (2014) and Covarrubias and Mendoza (2016) and the third is the definitive model resulting from this study. Table 2 shows the composition of items and dimensions.

Table 2
Estimated measurement models (items and latent factors).

	Original	Covarrubias	Proposed model
Instructional strategies	75	75	75
	76	76	76
	77	77	77
	78	78	78
	80		
	82		82
	84		
	86		86
Classroom management			70
	74	74	74
	79	79	79
	81	81	81
	83	83	83
	85	85	85
Student engagement	70	70	
	71	71	71
	72	72	72
	73	73	73
			80
			84
Efficacy in attention to students' singularity		80	
		82	
		84	
		86	

The principal variations with regards to the original relate to the fourth dimension of the structure of Covarrubias (2014) and Covarrubias and Mendoza (2016), centred on attention to students' singularities using alternative teaching strategies (item 80), adapting the level (item 82), using assessment strategies other than the usual ones (item 84) and providing appropriate challenges for the most capable students (item 86). In the original three-dimension model, these four items are found in the efficacy in instructional strategies factor.

The use of alternative teaching and evaluation strategies (items 80 and 84) in the final proposed model are placed in the student engagement dimension, along with helping students to think critically (item 71), motivating students with low interest (item 72) and fostering creativity (item 73).

All three models put items 71, 72 and 74 in the student engagement factor. Furthermore, in both the original model and in that of Covarrubias (2014), this factor includes one item relating to communicating with the most difficult students (item 70). In contrast, in the final model, this item is situated in the factor regarding efficacy in classroom management, alongside controlling bad behaviour in your classroom (item 74), managing a student who behaves badly (item 79), preventing problem students from spoiling the class (81), managing students who defy the teacher (83) and following the rules of the centre in class (85).

The models are evaluated using the robust normalised chi-squared test (χ^2/df), for which values below 5 are considered to be acceptable and below 3 very good. The index based on *RMSEA* (Root Mean Square Error of Approximation) residuals was also used, in which values below .08 are regarded as acceptable and below .06 optimal. Finally, the *CFI* (Comparative Fit Index) and *TLI* (Tucker-Lewis Index) are included, in which values greater than .9 are classed as acceptable and above .95 as good. An acceptable fit in the combination of the *RMSEA* and *CFI* indexes is sufficient evidence of validity (Hu & Bentler, 1999).

The dimensionality of the construct is reinforced with the interpretation of the correlations of latent factors, as well as study of convergent validity, which is undertaken using two indicators:

1) the average variance extracted (*AVE*) or ratio between the sum of the squared standardised factor loadings and the total number of items of the dimension. Values greater than .5 indicate that the explained variance of each factor is greater than the measurement error or, in other words, that more than 50% of the variability of the responses with the set of items that comprise the factor is explained; and
2) the Composite Reliability (CR), calculated based on the factor loadings and the error variance, also known as the Omega coefficient (McDonald, 1999). In this case, values greater than .7 are interpreted as reflecting good internal consistency in the factors. Hair et al. (2010) explain that AVE values between .4 and .5 are acceptable if the CR values are also acceptable.

Finally, the modification indexes are studied to identify possible causes of poor fit and to construct the optimal final model.

The explained variance by the model of the responses to each item (R^2) is also included.

Furthermore, the invariance of the final model between and its parameters were tested between the undergraduate and master's students using a multigroup confirmatory factorial model following the recommendations of Brown (2015). In a first step, the final three-dimension model is estimated separately for the undergraduate and master's groups using the same fit criteria to evaluate its quality. Next, a series of nested models are fitted to test the invariance of the configuration (equal structure of items and dimensions), the invariance of the metric (equal factorial weights) and the invariance of the scale (equivalent intercept and thresholds between categories).

To analyse invariance between groups, the differences between the fit values of the model with more fixed parameters compared with the next one, which has more free parameters, are studied. Owing to the sensitivity of chi² to sample size and non-normal distributions, which can lead it to conclude there is a lack of invariance when there is a significant worsening of the fit, Cheung and Rensvold (2002) recommend as a criterion of a worsening in the *CFI* index of .01 or more to determine lack of invariance. In addition, Chen (2007) suggests combining the changes in *CFI* with those in the *RMSEA* index, where an increase equal to or greater than .015 is also evidence of a lack of invariance be-

tween groups. All of the analyses of the confirmatory study were done using the MPlus 8 software (Muthen & Muthen, 2017).

SPSS is used in the analysis of the concurrent validity of the items to calculate CHAID (Chi Automatic Interaction Detection) (Kass, 1980), which provides the significance of the differences of means through chi², and CART (Classification And Regression Trees) (Breiman et al., 1984), making it possible to put the items in a hierarchy according to their normalised importance in a supervised process. These are classification techniques that have been proven to be of use for validating educational measures (Álvarez & Asensio-Muñoz, 2020; Blanco et al., 2017). In this case, the 17 items from the scale of Covarrubias (2014) are used as predictors and, as criterion, a global teacher self-efficacy item that the surveyed students also answer using a 5-point Likert scale.

Results

Table 3 shows the univariate descriptive statistics of the set of items, along with the bivariate correlations of each item with the criterion variable of the second objective.

Table 3
Description of the items in the sample of trainee teachers and Spearman correlations with the criterion item.

How much (A: None; B: A little; C: Partly; D: Quite a bit; E: Very Much) do you feel capable of …	*N*	Mean	*SD*	Rho
1 communicating with the most difficult students? (70, 24, A1)	701	3.74	.784	.378**
2 helping your students to think critically? (71, 22, A2)	740	4.05	.725	.276**
3 motivating students who show little interest in the subject? (72, 19, A3)	741	4.11	.717	**.420****
4 fostering your students' creativity? (73, 23, A4)	741	4.11	.830	.302**
5 controlling bad behaviour in your classroom? (74, 9, C1)	741	3.74	.825	.359**
6 evaluating whether your students understand what you have explained? (75, 7, B2)	741	4.01	.757	.273**
7 devising good questions for your students? (76, 3, B3)	741	3.93	.766	.300**
8 giving additional explanations or examples when your students are confused? (77, 2, B4)	741	4.22	.739	.279**
9 answering difficult questions that your students ask? (78, 5, B1)	741	3.79	.784	**.351****
10 managing a student who behaves badly or is "disruptive"? (79, 11, C3)	739	3.69	.838	.365**
11 implementing alternative teaching strategies in your classroom? (80, 4, D3)	738	4.01	.829	.333**
12 preventing one or a few problem students from spoiling your class? (81, 13, C4)	741	3.57	.849	.321**
13 adapting your classes to the correct level for each of your students? (82, 6, D1)	741	3.87	.805	.367**
14 managing students who openly defy you in a given moment? (83, 14, C1)	739	3.63	.929	.334**
15 using assessment strategies other than the usual ones? (84, 1, D1)	741	3.93	.874	.248**
16 getting students to follow the rules of the centre in class? (85, 10, C2)	740	3.93	.735	.343**
17 providing appropriate challenges for the most capable students? (86, 8, D4)	739	4.01	.811	.306**
18 being a good teacher in future? (87, criterion)	737	4.36	.754	1**

Note. The order in the complete questionnaire, in the original scale and in the version of Covarrubias (2014) is given alongside the item in parentheses. Correlation significant at 1% (**)

The global item (87) has the highest mean for perceived self-efficacy, with a score of 4.36 out of 5, followed by item 77, which centres on the capacity to give additional explanations or examples when your students are confused. On the other side, the items where the sample indicates a lower mean self-efficacy relate to the self-perceived capacity to: preventing one or a few problem students from spoiling your class (81), managing students who openly defy the teacher in a given moment (83) and managing a student who behaves badly (79).

The items that correlate most with the general perception of self-efficacy as a future teacher in the total sample are the capacity to motivate students who show little interest in the subject (72), with a rho value of .42, and the capacity to communicate with the most difficult students (70), with .38.

Proof of structural and convergent validity

The Mahalanobis distances did not identify any missing values or outliers. In contrast, Mardia's multivariate kurtosis measure indicates a lack of normality of the set of items from the attitude scale ($b2$ = 402.330; $N(b2)$ = 4.815; $p \leq$ .001). To do this, the matrix of polychoric correlations and robust estimators based on least squares (ULSMV) was used in the 3 models, the indexes of fit for which are shown in Table 4.

Table 4
Indexes of fit of the estimated confirmatory models.

Index	Original	Covarrubias	Proposed model
Chi2	662,742	596,597	476,418
df	116	113	113
p	< .001	< .001	< .001
Chi2/df	5,713	5,280	4,216
RMSEA	.083	.079	.068
RMSEA(LB 90%)	.077	.073	.062
RMSEA (UB 90%)	.089	.085	.075
RMSEA (p)	< .001	< .001	< .001
CFI	.924	.933	.950
TLI	.911	.920	.940

Considering the global fit that the normalised chi² index provides, an optimal value was not obtained in any of the cases (less than 3). The proposed model is the only one to achieve an acceptable value in this index (less than 5).

The root mean square error of approximation (*RMSEA*) has acceptable fit values (below .08) in the model of Covarrubias (2014) and in the proposed model, although in the former the upper limit is slightly higher (.085). The *CFI* and *TLI* comparative fit indexes are above .9 in the three models tested, reaching an optimal value in the proposed model (.95).

The study of convergent validity in the three models confirms that the factors are well defined, as Table 5 shows.

Table 5
Convergent validity indexes (AVE and Reliability).

Factors	Original	Covarrubias	Proposed Model
		AVE	
Instructional strategies	.387	.475	.417
Classroom management	.529	.529	.487
Student engagement	.398	.398	.406
Efficacy in attention to singularity		.400	
		Composite Reliability (CR)	
Instructional strategies	.833	.782	.809
Classroom management	.848	.848	.850
Student engagement	.724	.724	.773
Efficacy in attention to singularity		.726	

The *AVE* index is only above .5 in the classroom management dimension in the original model and in that of Covarrubias (2014), although both of these models have factors with *AVE* values slightly below .4. In the proposed model, all of the factors achieve a value greater than .4, which, combined with a CR close to .8, indicates acceptable convergent validity.

The best explained factor in the models is classroom management, with 50% approximately of variance explained in the proposed model and a CR of .85. As a group, the three models explain a similar variance in the answers to each of the items, as Table 6 shows, and in all cases it is .3 or greater. The items represented best in the models are 78 (answering difficult questions that your students ask), 79 (managing a student who behaves badly or is "disruptive") and 81 (preventing one or a few problem students from spoiling a class), with approximately 60% of explained variance in the different models.

Table 6
Explained variance for each item and total explained variance

	R2		
	Original	Covarrubias	Proposed model
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

The multigroup analysis shows the equivalence of the proposed model with the undergraduate and master's students (Table 7).

Table 7
Indexes of fit of the multigroup confirmatory factor analysis

	Undergraduate	Master's	Config.	Metric	Scalar	Metric vs Config.	Scalar vs Config.	Scalar vs Metric
Chi²	236,017	245,576	469,992	453,807	553,303			
df	113	113	226	240	271			
p	< .001	< .001	< .001	< .001	< .001			
Chi2/df	2,089	2,173	2,080	1,891	2,042			
RMSEA	.047	.075	.054	.049	.053			
RMSEA (LB)	.039	.063	.047	.042	.047			
RMSEA (UB)	.056	.088	.061	.056	.059			
RMSEA (p)	.677	< .001	.163	.576	.206			
CFI	.969	.908	.947	.953	.939			
Δ*CFI*						.006	-.009	-.014
Δ*RMSEA*						-.005	-.001	.004

The separate CFAs for the groups (undergraduate vs master's) have acceptable fit values (*RMSEA* < .08 and *CFI* > .9), although the fit is better with the sample of undergraduate students (*RMSEA* < .06 and *CFI* > .95).

The multigroup configuration model that does not place any restriction on the parameters of the groups, except for the structure of items and factors, displayed good *RMSEA* and *CFI* indices of fit. The model that fixes the factor loadings to keep them equal between the groups (metric) slightly improves the fit of the configuration model, although the changes are minimal (an improvement of approximately .005 in each index). Therefore, the invariance of the metric is demonstrated.

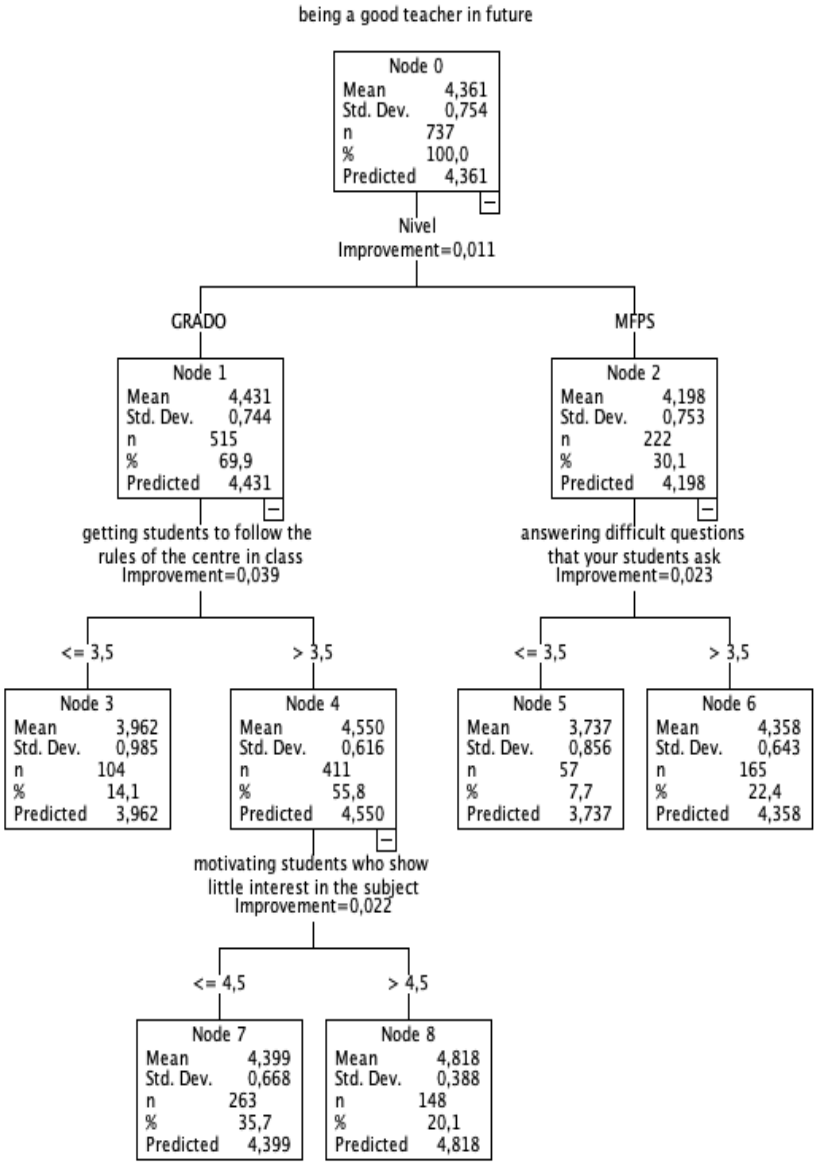
In the scalar model, where the intercepts and thresholds of the items remain equal between groups, a good fit is also achieved. Nonetheless, the changes in the fit with regards to

the model that tests the metric invariance become slightly worse, in particular *CFI*, which exceeds the recommended threshold of .01. The change in *RMSEA* is less than .005. Consequently, the scalar invariance cannot be ensured.

Evidence of concurrent validity

Taking as criterion the high inference item (87 in the complete questionnaire), items 72, 81, 82, 85 and 77 from the complete questionnaire emerge with CHAID as items that significantly discriminate between some of the nodes for the complete sample. The CART algorithm agrees that items 72 and 85 are the most important when it comes to explaining the overall TSE in the total sample. Figure 1, including as the first forced variable the level (undergraduate or master's), it can be seen that the undergraduate students with the lowest perceived teacher self-efficacy (M = 3.962), are situated in node 3: these are students who score 3 or less on item 85, relating to the capacity to get students to follow the rules. The undergraduate students with the best self-perceived overall self-efficacy (M = 4.818) are in node 8: they award themselves 4 or more on item 85 and 5 on item 72, that is to say, they feel they are highly capable of motivating students who show little interest. In the group of postgraduate students, the ones who feel most effective (M = 4.358) are in node 6 and are those who feel capable of answering difficult questions that students ask in class, that is to say, they have given themselves scores of 4 or more on item 78.

Figure 1
CART tree diagram for the total sample forcing the level as first variable.



In the subsample of undergraduate students, Figure 2 shows the relative importance of the variables in the explanation of the criterion according to the CART algorithm. In it, items 72 and 85, which also appear in the tree diagram in Figure 1, are followed by items 80, 82 and 86, which are significant in the CHAID tree shown in Figure 3 because they discriminate between nodes, taking 3 or 4 on the Likert scale as cut-off values as the case may be.

Figure 2
CART diagram of the importance of the variables in the subsample of undergraduate students.

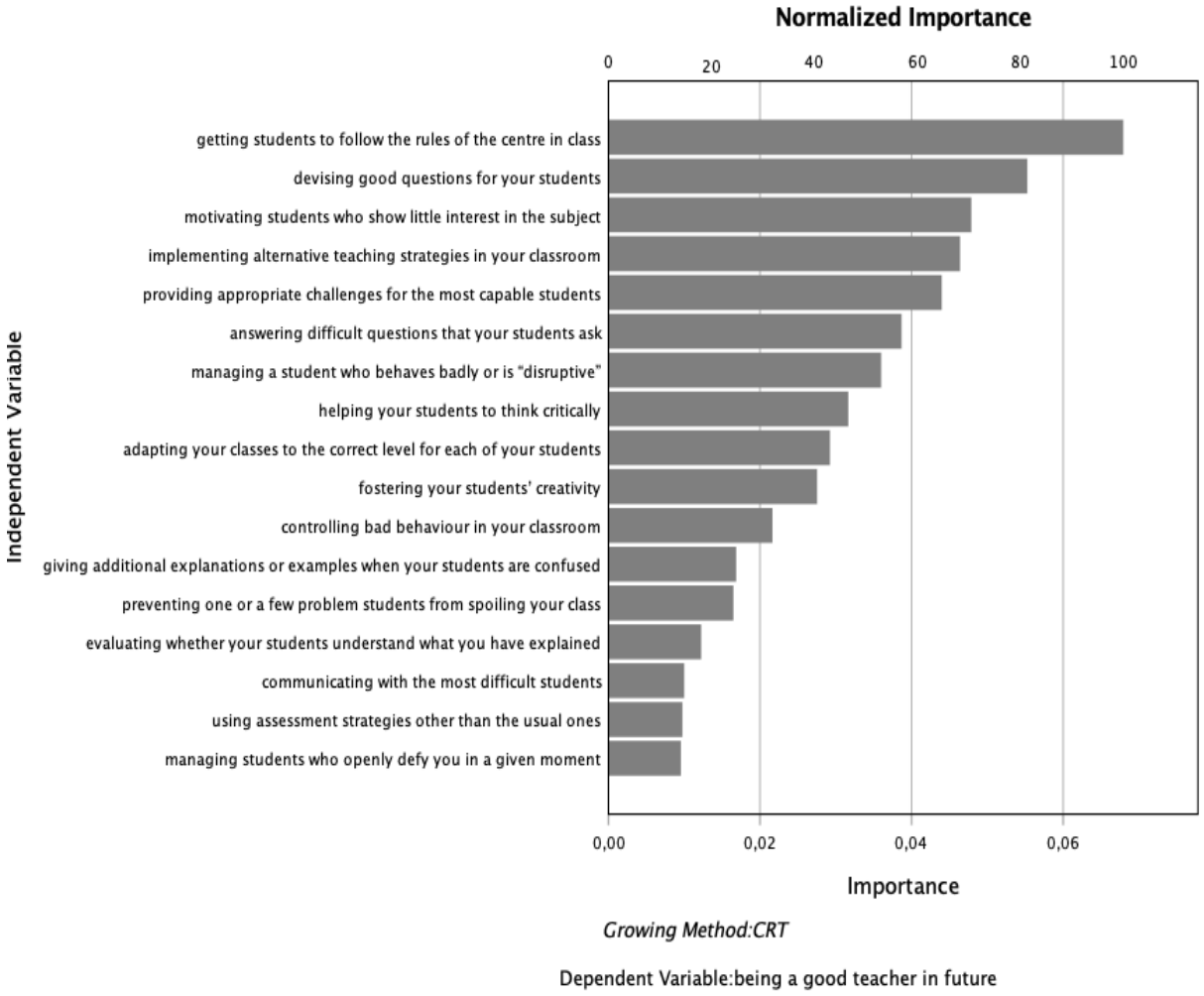
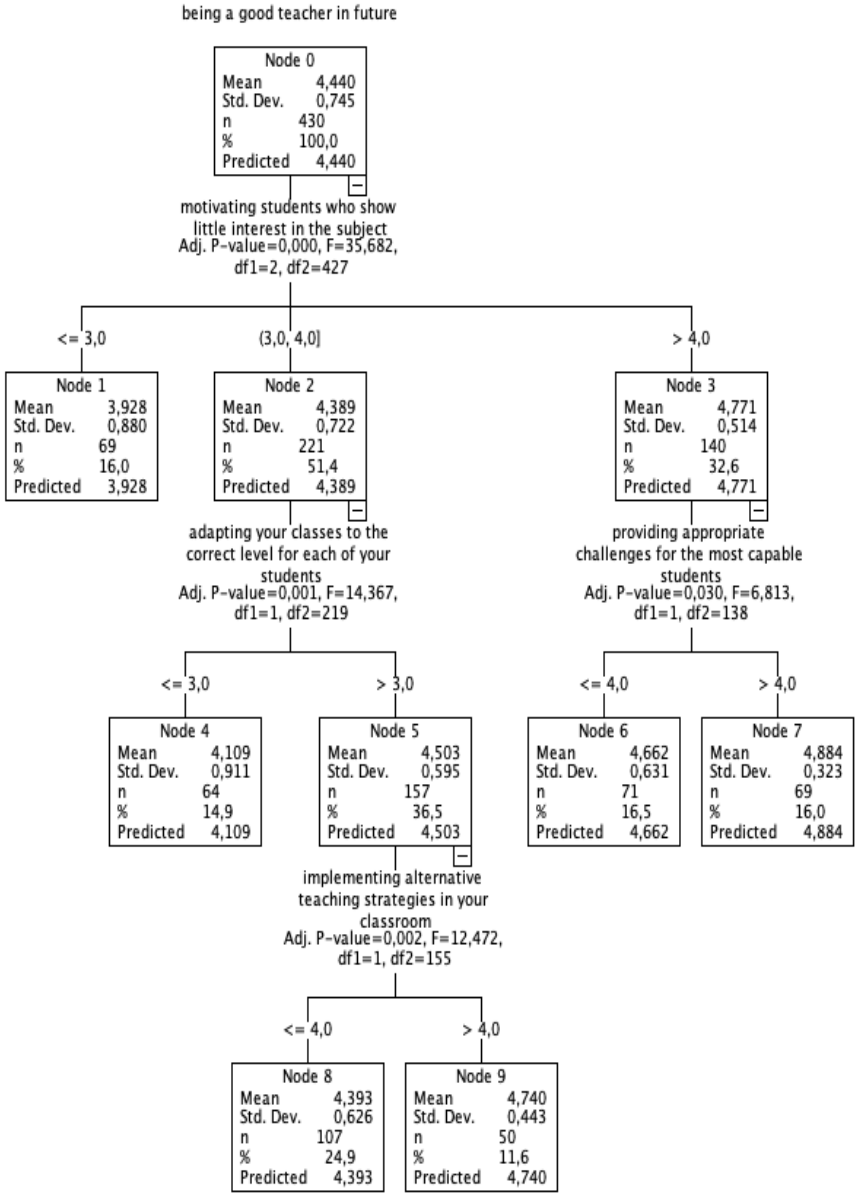


Figure 3
CHAID tree diagram in the subsample of undergraduate students.



Among the master's students, the order of importance of the variables changes, as Figure 4 shows, so that the most important variable in its relationship with the validity criterion used is clearly item 78, which refers to the capacity to answer difficult questions that students ask.

Figure 4
CART diagram of the importance of the variables in the subsample of postgraduate students.

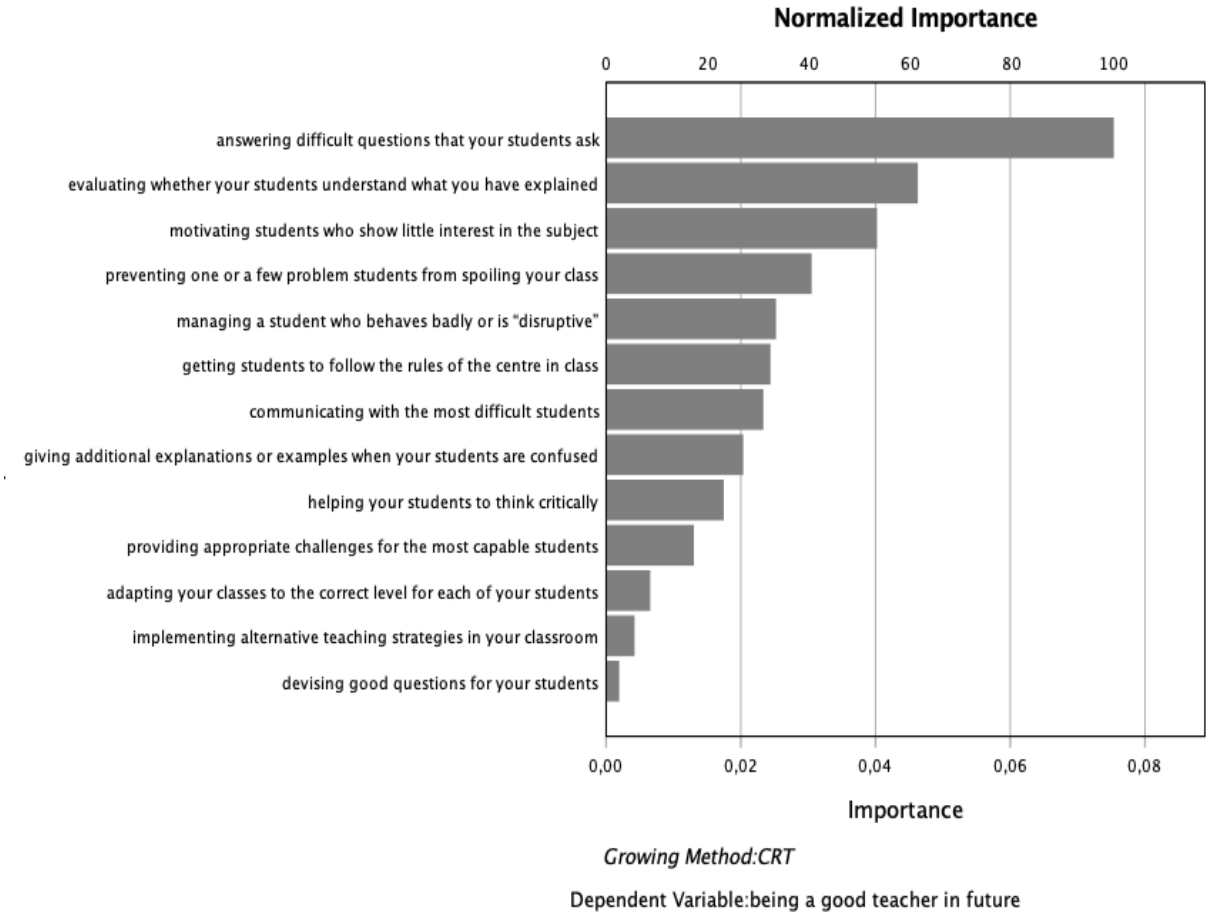
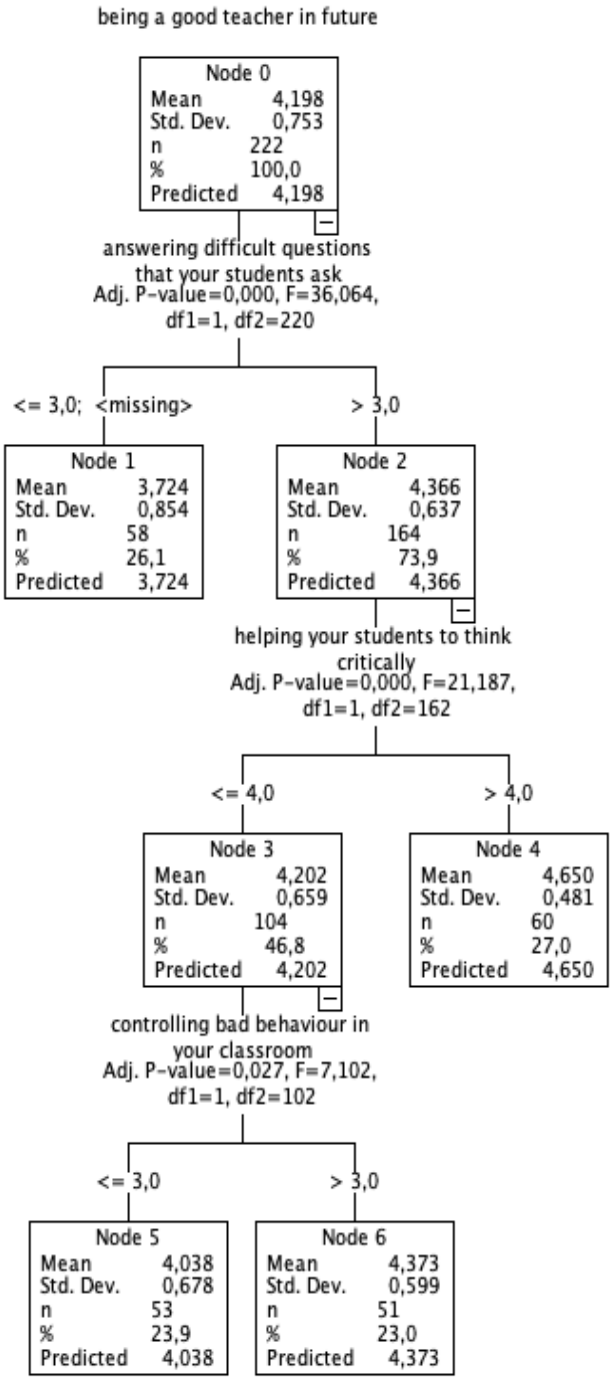


Figure 5 shows the tree for the surveyed master's students, which clearly shows that item 78 is the most important item for defining the concept of overall self-efficacy in this subgroup. In this tree diagram, it can be seen that a cut-off point of 4 in item 71 (capacity to help students think critically) makes it possible to divide node 2 and identify a subgroup of 60 students that make up node 4, in which the master's students with a higher perceived level of overall self-efficacy are found (M = 4.65).

Figure 5
CHAID tree diagram in the subsample of master's students.



Discussion

The results of the validation found in this study support a construct structured in three factors, which coincides with the one originally proposed and most of the research that has used this measurement model. Nonetheless, some items do not respond clearly to the original structure or the one proposed by Covarrubias (2014). On the one hand, perceived self-efficacy in the use of teaching and evaluation strategies, which in the original questionnaire form part of the handling of instructional strategies (items 80 and 84), are an observed expression of a new latent variable in the sample of inservice teachers from Chile. In the sample of students from Spain, the model improves when observed variables linked to the factor of student engagement are considered, which can be justified because using innovative strategies is normally associated with fostering motivation. The adaptive strategies (items 82 and 86), which Covarrubias takes from the original factor to create a fourth factor of attention to diversity, in the model proposed in this investigation are satisfactorily associated with the original dimension. Finally, item 70 (communicating with the most difficult students), in both the original model and in that of Covarrubias (2014) forms part of the student engagement factor, while in the model reached in the present study it is part of the classroom management dimension, which makes theoretical sense and can be explained by the translation of the main verb, which in English has several meanings.

On this point, it is important to note that the problem that translation into a language other than the original one causes in validation studies is in addition to the slippery or unclear nature of the construct. Covarrubias (2014), Covarrubias and Mendoza (2016) and Burgueño et al. (2019) use the back translation process to create the Spanish version of the TSES, while other researchers who have used this instrument with Spanish speakers do not specify the procedure for adapting it.

Moreover, measurement of the construct in Spanish may be affected by contextual variables such as the country (Chile, Colombia, Spain, Dominican Republic and Peru) or the type of teacher. Regarding the country of the sample, as well as the differences pertaining to each nationality, it is necessary to consider that the use of Spanish as a language differs in each one. As for the type of teacher, the validation works done with Spanish-speaking samples have mainly been done with inservice teachers. Only Burgueño et al. (2019) work with trainee teachers from a single training model. This study works with future teachers at a Spanish university from two models of teacher training: 1) students who will become early years or primary teachers (from 0 to 12 years) who are training on the relevant degrees, and 2) students who in one year will be qualified to give classes in secondary education, the baccalaureate and professional training, who are training on a master's (MSTT) accessed from a variety of university degrees in which they previously studied the disciplinary element. Accordingly, the work presented here considers groups not previously studied.

With regards to criterion validity, all of the items included in the analysis are relevant for predicting the included high-inference global item, general self-efficacy, understood as self-perception of the capacity to become a good teacher,

given that all of the correlations are significant, as Table 3 shows. The decision trees qualify this information and reveal that the items most linked to the type of perceived teacher self-efficacy of the undergraduate students, who are future EY and PT teachers, are the capacity to get students to follow the rules in class (85) and the capacity they feel they have to motivate students who show little interest in the subject (72). In contrast, at the postgraduate level, the most relevant item is 78, which relates the capacity they feel they have of answering difficult questions.

Conclusions

This study achieves the established objectives as acceptable indexes of fit are obtained for the dimensionality of the proposed teacher self-efficacy and a consistent and valid model is achieved in both of the subgroups considered, undergraduate students and master's students. Furthermore, evidence is found for the utility of the items for predicting the students' perception of their global teacher self-efficacy. One interesting contribution is the differential importance of certain items in explaining the criterion students choose according to whether they are undergraduates or master's students. Accordingly, it underscores the value of knowledge of the subject to be taught in the self-efficacy of the master's students, who are mainly future secondary teachers, while with the undergraduate students, who are future teachers of children aged from 0 to 12, rules and motivation have the greatest weight in their TSE. Although the construct responds to the same factorial structure, as shown in the multigroup CFA, the undergraduate and master's students have differing self-efficacy profiles and the items in the questionnaire help discriminate between them.

These conclusions comprise an original contribution by this study, which also features the use of an analysis methodology specifically fitted to the intended objectives and the characteristics of the data.

The principal limitations of the research presented include those deriving from the characteristics of the sample and the sampling procedure, which reduce the capacity for generalisation of the results found. In consequence, more primary research is necessary into the topic to provide evidence about the stability of the construct in contexts other than the one used here.

Despite this, the study done, with future teachers at the moment that they start their training, has interesting practical implications regarding the most relevant training actions in each case. considering the differing profiles identified. In future research, it would also be of interest to study the evolution of the construct throughout the degree programmes and the effects that specific intervention programmes adapted to different groups might have.

Conflict of interest.- The authors of this article declare no conflict of interest.

Financial support.- No funding.

References

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-Casaubón, 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

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/3FTUb0r

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

Rodriguez, 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

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/1.3102/00346543156268O1