Received: 6 December 2022 | Accepted: 29 May 2023

DOI: 10.1111/1471-3802.12607

ORIGINAL ARTICLE

Development and validation of a short form of the Teacher Efficacy for Inclusive Practices Scale (TEIP-SF)

C. Sahli Lozano[1] | S. Wüthrich[1] | N. Baumli[1] | U. Sharma[2] | T. Loreman[3] | C. Forlin[4]

[1]Institute for Research, Development and Evaluation, Bern University of Teacher Education, Bern, Switzerland

[2]Faculty of Education, Monash University, Melbourne, Victoria, Australia

[3]Faculty of Education, Concordia University of Edmonton, Edmonton, Alberta, Canada

[4]Notre Dame University, Perth, Western Australia, Australia

Correspondence
Caroline Sahli Lozano, Pädagogische Hochschule Bern, Institut für Forschung, Entwicklung und Evaluation, Fabrikstrasse 8, 3012 Bern, Switzerland.
Email: caroline.sahlilozano@phbern.ch

Abstract

High self-efficacy is a marker of successful teaching and is, therefore, a subject of great interest to research on inclusive education. One of the most frequently used instruments to assess such beliefs is the Teacher Efficacy for Inclusive Practice (TEIP) scale. Although used widely, some studies did not precisely replicate the original factor structure, and no short form of the TEIP scale currently exists, although this could enhance measurement efficiency. This study (1) systematically assessed the TEIP scale's factor structure and psychometric properties, (2) identified potentially problematic items and developed a more concise short form of the scale, and (3) evaluated its dimensionality and criterion and convergent validities using three validation samples of teachers in three different countries (486 in Switzerland, 189 in Australia and 276 in Canada). Compared to the full-length TEIP scale, the TEIP-SF uses half the items, demonstrates better model fit and reveals a clearer distinction of domain-specific factors. In conclusion, the TEIP-SF represents a concise, efficient means of assessing teachers' self-efficacy about teaching in inclusive classrooms.

KEYWORDS
confirmatory factor analysis, inclusive education, measurement invariance, measurement scale, teacher self-efficacy, TEIP

Key Points

- The Teacher Efficacy for Inclusive Practice (TEIP) scale is one of the most prominent scales to measure teachers' self-efficacy regarding inclusive education.
- Using teacher samples from Switzerland, Australia and Canada, the current study developed and validated a more concise short form of the TEIP scale (TEIP-SF).
- The TEIP-SF is more time efficient than the original scale, sufficiently valid and reliable, and demonstrates a robust factor structure across international teacher samples.

INTRODUCTION

Teacher self-efficacy and inclusion

Teacher self-efficacy (TSE) refers to the subjective belief that one can promote student learning outcomes even when conditions are difficult and challenging (Bandura, 1977; Tschannen-Moran & Hoy, 2001). It has become a remarkably popular indicator in recent decades because of its positive relationships with teaching effectiveness, student outcomes, teacher well-being and job satisfaction (Zee & Koomen, 2016). Yet TSE is

a multidimensional construct made up of different domains. For example, the well-validated Teacher's Sense of Efficacy Scale (TSES) examines the three factors of *efficacy for instructional strategies*, *efficacy for classroom management*, and *efficacy for student engagement* (Tschannen-Moran & Hoy, 2001). In contrast, other scales identify additional factors, such as *cooperation with colleagues and parents* or *coping with changes* (Skaalvik & Skaalvik, 2007).

Teacher self-efficacy, therefore, is an important concept in research on inclusive education. Having to teach students with a range of abilities and support needs creates additional challenges for teachers (Pit-ten Cate et al., 2019). Efforts have been made to adapt existing scales specifically to measure TSE in inclusive classrooms. The most prominent scale in this regard is the Teacher Efficacy for Inclusive Practices (TEIP) scale (Sharma et al., 2012).

The TEIP scale and its factor structure

The TEIP scale was developed from existing scales measuring general TSE (e.g. the TSES), a literature review and advice from experts in the field of inclusive education (Sharma et al., 2012). An exploratory factor analysis (EFA) based on data concerning pre-service teachers from Canada, Australia, Hong Kong and India identified three factors: (1) *Efficacy in using Inclusive Instructions* (*EI*, i.e. the ability to use teaching strategies that promote the inclusion of all learners), (2) *Efficacy in Managing behaviour* (*EM*; i.e. the ability to prevent or cope with disruptive student behaviour) and (3) *Efficacy in Collaboration* (*EC*; i.e. the ability to work with parents and other professionals; Sharma et al., 2012).

Within 10 years of its publication, the TEIP scale had been used in at least 95 studies and translated into more than 13 languages. The structure of three correlated factors was replicated for various countries and teacher samples (e.g. Alnahdi, 2019; Cardona-Molto et al., 2020; Narkun & Smogorzewska, 2019; Savolainen et al., 2012). Often, however, slight modifications (e.g. removing items) were necessary to reach an acceptable model fit in confirmatory factor analyses (see Table 1 for examples of scale modification), but this has limited the ability to make comparisons across study results.

A second issue was the TEIP scale's 18-item length, especially when considering complex research questions requiring the collection of data about multiple other constructs. Advantages of shorter scales are their efficiency (less collection of redundant information) and a smaller risk that participants will answer carelessly (Gibson & Bowling, 2020). The TSES, for example, also exists in a 12-item short form (Tschannen-Moran & Hoy, 2001).

Thus, developing a short form of the scale, by discarding items of limited relevance, should help to improve measurement consistency and scale efficiency. The present study (1) systematically assessed the TEIP scale's factor structure and psychometric properties, (2) identified potentially obsolete or cross-loading items and developed a more concise, short form of the TEIP scale to maximize its efficiency, and (3) evaluated that short form's dimensionality using three validation samples from three different countries and assessed its criterion and convergent validity.

METHOD

Participants

A sample of teachers was recruited via a link in an online survey sent out to a regional professional association of teachers (around 10,000 members in the canton of Bern) and two hundred randomly chosen schools in other German-speaking parts of Switzerland. A total of $N = 1557$ teachers participated in the survey. Only surveys completed by regular, in-service preschool, primary and secondary teachers ($n = 1035$) were considered for analysis. After removing 62 cases identified as multivariate outliers (Osborne, 2014), the $n = 973$ participants remaining were split into a test (Sample 1) and an independent validation sample (Sample 2; see Section 2.3). The sample characteristics of teaching level, sex, and age were comparable to those of Switzerland's general teacher population (Swiss Federal Statistical Office, 2018; see Table 2). We used an Australian dataset of 149 in-service secondary school teachers and a Canadian dataset of 276 in-service teachers (preschool, primary and secondary school) as further validation samples (Samples 3 and 4, respectively). Teachers in Australia were recruited via an online invitation to participate in the study sent to 613 school principals in the state of Victoria; 12 schools agreed to participate and sent the link to their teachers. With the assistance of their school districts, Canadian teachers from rural and urban areas in the provinces of Alberta and British Columbia were emailed a link to a voluntary online survey.

Instrumentation

Participants evaluated the TEIP scale's 18 statements using 6-point Likert scales ranging from 1 (strongly disagree) to 6 (strongly agree). An updated version of the original scale was used where Item 14, 'I can make parents feel comfortable coming to school', had been replaced by 'I can improve the learning of a student who is failing' (both items are listed in the subfactor of *Efficacy in Collaboration*). Participants in Switzerland used the German translation of this scale by Gebhardt et al. (2018).

Procedure and statistical analyses

All data preparation and analyses were done using the *R* statistical programming language (R Core Team, 2019).

TABLE 1 Overview of studies assessing the TEIP scale's structure using Confirmatory Factor Analysis (CFA).

Author/Year	Sample size/Sample type	Adaptions	Model fit	Factor reliability	Model used
Aiello et al. (2018)	Total $N=352$ Italy Pre- and in-service teachers	Removed items: 18, 11, 2 (16, 17, 18)	$\chi^2=683.5$ CFI = 0.918 SRMR = 0.060 RMSEA = 0.049	–	Correlated-factors model, in combination with the SACIE-R
Alnahdi (2019)	Total $N=432$ Saudi Arabia Pre- and in-service teachers	Removed items: 11, 12 (1, 17) Covary errors terms	$\chi^2=308.8$ CFI = 0.935 SRMR = 0.043 RMSEA = 0.066	$\alpha_{total}=0.93$ $\alpha_{EI}=0.82$ $\alpha_{EM}=0.83$ $\alpha_{EC}=0.82$	Correlated-factors model
Cardona-Molto et al. (2020)	Total $N=475$ Spain Pre-service teachers	Removed items: 1, 12, 13, 14, 18 (1, 12, 13, 14, 19)	$\chi^2=429.5$ CFI = 0.864 RMSEA = 0.090	$\alpha_{total}=0.92$ $\alpha_{EI}=0.84$ $\alpha_{EM}=0.89$ $\alpha_{EC}=0.85$	Correlated-factors model
Chao et al. (2016)	Total $N=417$ Hong Kong In-service teachers	Removed items: 11, 3 (8, 10)	$\chi^2=363.3$ CFI = 0.90 SRMR = 0.060 RMSEA = 0.08	α values between 0.78 and 0.82	Correlated-factors model
Malinen, Savolainen, Engelbrecht, et al. (2013)	Total $N=1911$ China = 451 Finland = 855 South Africa = 605 In-service teachers	China–Removed items: 3, 13 Finland—Removed items: 12, 18 South Africa—Removed items: 3	$\chi^2=272.0–456.5$ CFI = 0.92–0.94 SRMR = 0.05–0.06 RMSEA = 0.06–0.06	$\alpha_{total}=0.90–0.91$ $\alpha_{EI}=0.75–0.77$ $\alpha_{EM}=0.85–0.88$ $\alpha_{EC}=0.83–0.87$	Correlated-factors model tested for each country
Malinen et al. (2012)	Total $N=451$ China In-service teachers	Removed items: 3, 13	$\chi^2=272.0$ CFI = 0.92 SRMR = 0.06 RMSEA = 0.06	$\alpha_{total}=0.91$ $\alpha_{EI}=0.75$ $\alpha_{EM}=0.89$ $\alpha_{EC}=0.88$	Correlated-factors model
Malinen, Savolainen, and Xu (2013)	Total $N=552$ China Pre-service teachers	Removed item: 12 Covariation of errors terms	$\chi^2=3703.7/3703.7$ CFI = 0.96/0.96 SRMR = 0.04/0.04 RMSEA = 0.05/0.05	$\alpha_{total}=0.90$ $\alpha_{subscales}=0.75–0.85$	Correlated-factors model/hierarchical model
Miesera et al. (2019)	Total $N=909$ Germany Pre-service teachers	Removed items: 6, 14, 17	CFI = 0.913 SRMR = 0.069 RMSEA = 0.066	–	Correlated-factors model
Mohamed Emam and Al-Mahdy (2020)	Total $N=287$ Sultanate of Oman In-service teachers	Removed items: 3, 7, 13 Covariation of errors terms	$\chi^2=238.2/238.9$ CFI = 0.961/0.961 RMSEA = 0.079/0.079	$\alpha_{total}=0.93$ $\alpha_{EI}=0.83$ $\alpha_{EM}=0.93$ $\alpha_{EC}=0.93$	Correlated-factors model/hierarchical model
Opoku (2021)	Total $N=82$ Ghana In-service teachers	No adaptions were made	$\chi^2=232.614$ CFI = 0.90 RMSEA = 0.05	–	Correlated-factors model
Park et al. (2016)	Total $N=134$ USA Pre-service teachers	Removed items: 6, 14 (6, 14)	$\chi^2=260.6/181.2$ CFI = 0.994/0.997 RMSEA = 0.085/0.064	$\alpha_{total}=0.98$ $\alpha_{EI}=0.93$ $\alpha_{EM}=0.94$ $\alpha_{EC}=0.95$	Correlated-factors model/bi-factor model
Sharma and Jacobs (2016)	Total $N=602$ Australia = 253 India = 349 In-service teachers	Australia—Removed items: 11, 12, 18 India—Removed item: 5	$\chi^2=9.53–16.72$ CFI = 0.99–1.00 RMSEA = 0.00–0.05 SRMR = 0.02–0.04		Correlated-factors model, in combination with AIS and ITICS scale, using a separate CFA for each country

(Continues)

TABLE 1 (Continued)

Author/Year	Sample size/Sample type	Adaptions	Model fit	Factor reliability	Model used
Tanrıverdi and Özokçu (2018)	Total $N=567$ Turkey Pre-service teachers	No adaptions were made	$\chi^2=887.1$ CFI = 0.96 RMSEA = 0.101 SRMR = 0.054	$\alpha_{total}=0.89$ $\alpha_{EI}=0.77$ $\alpha_{EM}=0.68$ $\alpha_{EC}=0.79$	Correlated-factors model
Vogiatzi et al. (2022)	Total $N=465$ Greece In-service teachers	No adaptions were made	$\chi^2=77.84$ CFI = 0.93 RMSEA = 0.013 SRMR = 0.003	$\alpha_{total}=0.89$	Correlated-factors model
Yada et al. (2019)	Total $N=1384$ Finland = 1123 Japan = 261 In-service teachers	Covariation of errors terms	$\chi^2=2461.6/2667.7$ CFI = 0.933/0.925 RMSEA = 0.046/0.048 SRMR = 0.072/0.075	–	Correlated-factors model/hierarchical model in combination with STSE scale with restricted loadings and intercepts, using a multi-group CFA
Yada et al. (2021)	Total $N=105$ Finland Pre-service teachers	Covariation of errors terms	CFI = 0.948 RMSEA = 0.060 SRMR = 0.061	$\alpha_{total}=0.93$	Correlated-factors model

Note: Item numbers in brackets refer to the item numbering in the respective studies.
Abbreviations: EC, efficacy in collaboration; EI, efficacy to use inclusive instructions; EM, efficacy in managing behaviour.

TABLE 2 Participants' backgrounds.

Variable	Sample 1 (Switzerland) *n* (%)	Sample 2 (Switzerland) *n* (%)	Sample 3 (Australia) *n* (%)	Sample 4 (Canada) *n* (%)
Sex				
Female	381 (78.2)	380 (78.2)	94 (63.1)	208 (75.4)
Male	103 (21.1)	100 (20.6)	52 (34.9)	56 (20.3)
Other	3 (0.6)	6 (1.2)	0 (0)	12 (4.3)
Missing	0 (0)	0 (0)	3 (2.0)	0 (0)
Age				
<25 years	22 (4.5)	21 (4.3)	8 (5.4)	10 (3.6)
25–30 years	73 (15.0)	72 (14.8)	25 (18.8)	37 (13.4)
31–40 years	94 (19.3)	93 (19.1)	41 (27.5)	89 (32.2)
>40 years	298 (61.2)	300 (61.7)	72 (48.3)	136 (49.3)
Missing	0 (0)	0 (0)	0 (0)	4 (1.4)
Teaching experience				
<1 year	6 (1.2)	5 (1.0)	2 (1.3)	N/A
1–3 years	49 (10.1)	46 (9.5)	33 (22.1)	N/A
4–10 years	91 (18.7)	92 (18.9)	44 (29.5)	N/A
>10 years	341 (70.0)	343 (70.6)	64 (43.0)	N/A
Missing	0 (0)	0 (0)	6 (4.0)	N/A
Teaching level				
Preschool	102 (20.9)	101 (20.8)	0 (0.0)	7 (2.5)
Primary	272 (55.6)	273 (56.2)	0 (0.0)	166 (60.1)
Secondary	113 (23.2)	112 (23.0)	149 (100.0)	94 (34.1)
Missing	0 (0)	0 (0)	0 (0)	9 (3.3)
Sample size (*N*)	487	486	149	276

Note: The teaching experience variable was not collected/unavailable (N/A) for analysis in the Canadian sample.

For data preparation, the Swiss sample was split into two subsamples using the *Anticlust* package (Papenberg & Klau, 2021)—one subsample for testing (Sample 1) and one for validation (Sample 2). *Anticlust* forms groups so that between-group similarities and within-group heterogeneity are maximized. These (dis)similarities are measured using the pairwise distances in a predefined feature matrix, which, in this case, included the variables of sex, age, teaching experience and teaching level. The resulting subsamples demonstrated equally distributed background variables (see Table 2).

Statistical analyses followed three steps. In step (1), confirmatory factor analyses (CFAs) were calculated based on the original model with three correlated factors (Sharma et al., 2012), and fit indices and item-factor loadings were used to evaluate the resulting models and assess factor reliabilities. In step (2), additional EFAs were applied to examine the TEIP scale's dimensionality and identify potential sources of model–data misfit (Schmitt et al., 2018). Based on these results, a short form of the original scale was developed by omitting items with problematic content and factor loadings. In step (3), the short form's dimensionality was evaluated, the reliability, criterion validity, and convergent validity of the short and original TEIP scale versions were compared, and measurement invariance across samples was assessed. In the next sections, these steps are outlined in more detail.

Step 1: Assessment of the TEIP scale's factor structure and reliability

All analyses were made using the *lavaan* package (Rosseel, 2012). Maximum likelihood estimation, with robust (Huber–White) standard errors and a scaled test statistic (estimator: 'MLR'), was used to estimate the CFA models due to the data's non-normal distribution.

Model–data fit was assessed using the following fit indices: (1) the chi-square test, (2) the comparative fit index (CFI), (3) the standardized root mean square residual (SRMR) and (4) the root mean square error of approximation (RMSEA). Because the chi-square test is sensitive to sample size (Dimitrov, 2012), our focus was predominantly on CFI, SRMR and RMSEA. Model fit was considered acceptable with CFI >0.90, SRMR <0.08 and RMSEA <0.08, and good with CFI >0.95, SRMR <0.05 and RMSEA <0.05 (Hu & Bentler, 1999).

Additionally, the conventional Cronbach's alpha (α) and the more recommended omega (ω) were calculated as measures of global (i.e. unadjusted) factor reliability. Omega's advantages over α are independence in the number of items used and the lack of assumptions regarding essential tau-equivalence (Hayes & Coutts, 2020; McNeish, 2018). Omega's value can be interpreted similarly to alpha's (i.e. >0.7 acceptable, >0.8 good and >0.9 excellent reliability).

Step 2: Creation of a short form (TEIP-SF)

In cases of poor model fit or poor psychometric properties following a CFA, calculating an EFA is recommended to find out where that model misfit occurs, even if there are strong theoretical reasons for a given factor's structure (Schmitt et al., 2018). To decide how many factors to extract for each EFA, we considered the scree plot, the parallel analysis and the minimal average partial (MAP) as criteria together with theoretical considerations (Howard, 2016). The following statistical packages were used for the analyses: *lavaan* (Rosseel et al., 2022) for the EFAs (with the MLR estimator as the extraction method and the direct oblimin as the rotation method), *psych* (Revelle, 2022) for the parallel analysis and *EFA.dimensions* (O'Connor, 2022) for the MAP.

Based on the item-factor loading patterns, modifications to the original scale were proposed by removing items with the lowest factor loadings and/or highest cross-loadings, with the goal of achieving a simple structure of the TEIP scale and a more efficient short form. Decisions on which items to retain or remove were based on three criteria (Howard, 2016; Watkins, 2018): (1) substantial factor loading (omitting items if they had a weak standardized factor loading on the latent factors, <0.4), (2) item–content fit (omitting items if their content did not fully match the subfactor's narrative) and (3) low cross-loadings (omitting items with substantial loadings on other factors, >0.3).

Step 3: Evaluation of the TEIP-SF's reliability and validity

The shortened TEIP scale's (TEIP-SF) dimensionality (using CFAs of the correlated-factor model) and factor reliabilities were evaluated using the additional validation samples of teachers from Switzerland (Sample 2), Australia (Sample 3) and Canada (Sample 4). Criterion and convergent validity analyses of the TEIP-SF included unweighted scale score correlations between the TEIP scale, the TEIP-SF and the Attitudes to Inclusion Scale (AIS; Sharma & Jacobs, 2016) as well as correlations with a single item assessing teachers' overall self-efficacy about the inclusion of students with special educational needs ('Please rate your level of confidence in teaching students with a disability in a regular classroom'). The AIS measures teachers' attitudes towards the concept of inclusive education via eight items and two factors (*Beliefs* and *Feelings*). Because attitudes and self-efficacy towards inclusive education are known to correlate moderately ($\bar{r}=0.35$; Yada et al., 2022), we took similar or higher correlations between the AIS and the TEIP scale/TEIP-SF's overall self-efficacy item as evidence of convergent validity. The TEIP-SF scale's efficiency was assessed by inspecting patterns of correlations, where similar or even higher correlations between the TEIP-SF and the AIS

and the overall self-efficacy item would indicate greater precision and/or efficiency. Finally, using multi-group confirmatory factor analysis (MGCFA), we assessed measurement invariance to determine whether the scale measures the same latent constructs in the same way across the three validations samples by comparing the configural, metric and scalar invariance models. While in metric invariance models, factor loadings between the items and the latent variable are constrained to be equal across samples, scalar invariance models also constrain the item intercepts. Demonstration of metric invariance allows for meaningful comparisons of relationships between latent constructs, while demonstration of scalar invariance allows for meaningful comparisons of latent means across samples (Cieciuch & Davidov, 2015). If full invariance could not be established, stepwise tests for partial invariance by freeing either factor loadings (for partial metric invariance) or intercepts (for partial scalar invariance) of non-invariant items were conducted. According to Cieciuch and Davidov (2015), the loadings and intercepts of at least two items per construct must be equal across groups to establish partial invariance, and according to Byrne et al. (1989) and Steenkamp and Baumgartner (1998), partial invariance is sufficient for meaningful cross-group comparisons. For the critical thresholds of model comparisons, Chen's (2007) recommendations for samples larger than 300 were used, where changes in RMSEA <0.015, in CFI <0.01 and SRMR <0.03 when moving from configural to metric invariance models and changes in RMSEA <0.015, CFI <0.01 and SRMR <0.01 when moving from metric to scalar invariance models would imply metric or scalar measurement invariance, respectively.

RESULTS

Step 1: The TEIP scale's factor structure and reliability

For Samples 3 and 4 (Australia and Canada), although the model fits for the full scale and the correlated-factors model with three factors were acceptable, the model yielded a poor fit for Sample 1 (Switzerland) with all the fit indices below acceptable thresholds (see Table 4). Modifications were needed to reach a good fit. Additional EFAs were thus calculated to see where model misfit occurred in Sample 1 and how the scale could be improved. Based on these results, a more concise, efficient short form of the TEIP scale was developed.

Step 2: Development of a short form (TEIP-SF)

To decide how many factors to extract for the EFAs, scree plot, parallel analysis and MAP criteria were considered, as were theoretical assumptions. Because the scree plot suggested two factors, the MAP criteria suggested three, and the parallel analysis suggested four (Figure 1), we ran EFAs with factor extractions from one to four factors.

While the three- and four-factor solutions revealed acceptable (three factors) to good (four factors) model fits, the two-factor and one-factor solutions did not (Table 3, step 2). Nevertheless, the two-factor solution yielded an interesting pattern: the EI and EC factors collapsed into one factor, whereas the EM factor remained independent (Table 4, step 2). Similarly, when comparing the three- and four-factor models, the EM factor's (F2) loading patterns were very robust, with all loadings ≥0.40 and only Item 12 having minimal cross-loadings (≥0.3) on other factors. The EI Factor (F1) had four items (1–4) that remained robust in the three- and four-factor models, but Items 5 and 6 had rather low loadings in the four-factor solution and low cross-loadings in both solutions. The most notable pattern was found for the EC subfactor in the three- and four-factor models. Here, Items 15–18 were robust in both solutions, but Items 13 and 14 did not load on the intended factor. In the four-factor solution, these two items even represented a new factor closely aligned to the *student engagement* factor in the TSES (Tschannen-Moran & Hoy, 2001) that guided the development of items for the TEIP.

General loading patterns supported the EI, EM and EC factors: the three-factor model's fit was acceptable-to-good, which accorded with the strong theoretical grounds for the TEIP scale's three-factor structure. The EFAs also suggested that some items could be discarded because of misfits in their content and factor cross-loadings. Loading patterns also suggested the feasibility of a short form of the scale, with advantages regarding factor conciseness and test economy. To this end, the three-factor solution was used as a starting point for developing a short form of the TEIP scale: the TEIP-SF.

Three criteria for selecting adequate items for these three factors were considered: (1) substantial factor loadings, (2) item contents and (3) minimal cross-loadings. To shorten the scale as much as possible while considering factor reliability, a minimum of three items per factor was retained. Table 5 gives an overview of the items chosen for the TEIP-SF.

For the *Efficacy to use Inclusive Instructions* factor, Items 1, 2 and 3 were retained as they had robust factor loadings. The decision to use Item 3 was difficult because of its frequent misfits identified in previous studies (e.g. Chao et al., 2016; Malinen, Savolainen, Engelbrecht, et al., 2013; Mohamed Emam & Al-Mahdy, 2020). However, keeping this item rather than Item 4 seemed to match the factor better in terms of content; otherwise, its meaning would have shifted towards mainly reflecting learning assessment.

For the *Efficacy in Managing Behaviour* factor, Items 7, 9 and 10 were retained. Items 7–10 all had very high loadings (>0.7). In terms of content, they all reflected the ability to prevent and control disruptive student

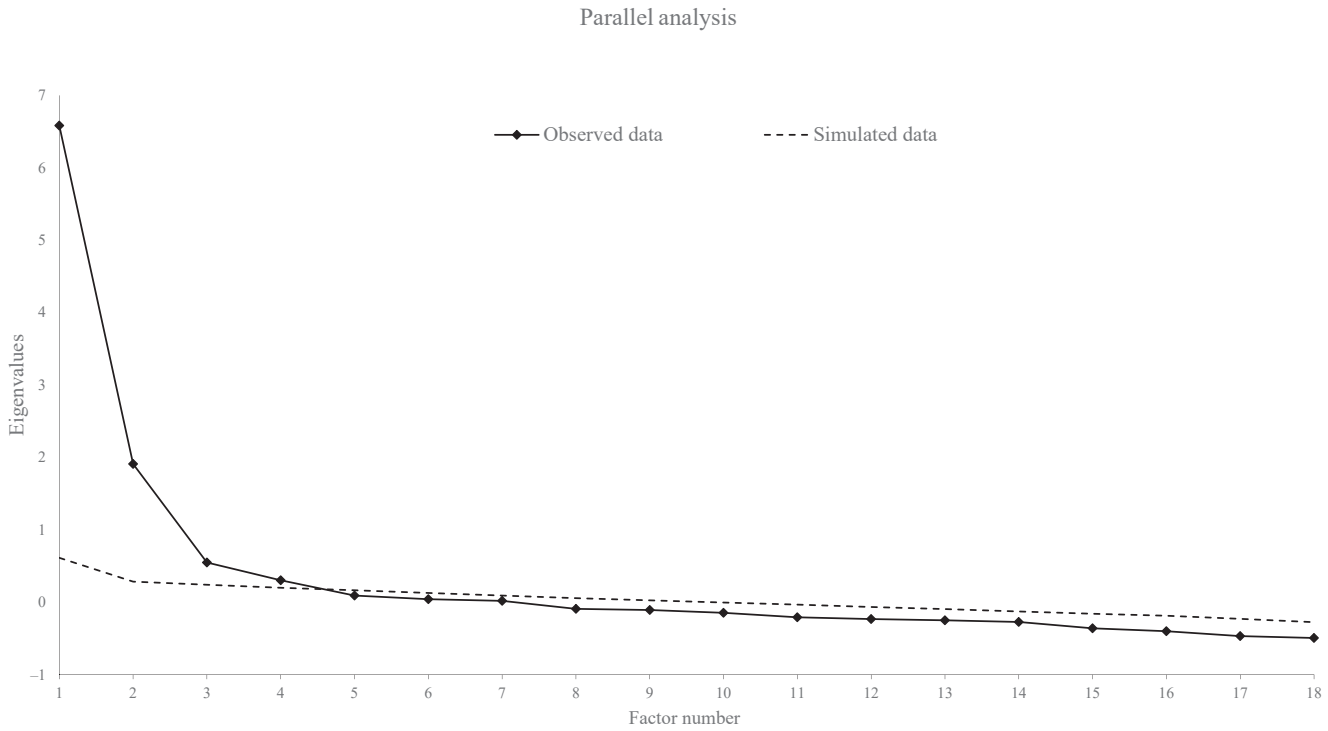

FIGURE 1 Parallel analysis of observed and simulated eigenvalues.

TABLE 3 CFA and EFA for the TEIP scale and TEIP-SF: model fits.

Sample	Model	χ^2	df	χ^2/df	CFI	SRMR	RMSEA (90% CI)
Step 1: Confirmatory factor analysis							
1	Correlated factors	722.640	132	5.47	0.873	0.096	0.102 (0.095–0.110)
3	Correlated factors	211.736	132	1.60	0.925	0.070	0.069 (0.051–0.086)
4	Correlated factors	256.535	132	1.94	0.941	0.050	0.071 (0.058–0.084)
Step 2: Exploratory factor analysis							
1	4 factors	157.405	87	1.809	0.985	0.017	0.043 (0.032–0.540)
1	3 factors	304.596	102	2.989	0.958	0.028	0.067 (0.058–0.076)
1	2 factors	591.619	118	5.014	0.899	0.047	0.096 (0.089–0.104)
1	1 factor	1605.463	135	11.892	0.679	0.116	0.161 (0.154–0.168)
Step 3: Evaluation of reliability and validity TEIP-SF							
2	Correlated factors	52.688	24	2.20	0.985	0.035	0.054 (0.034–0.073)
3	Correlated factors	22.187	24	0.92	1.000	0.038	0.000 (0.000–0.065)
4	Correlated factors	33.296	24	1.39	0.989	0.038	0.046 (0.000–0.081)

Note: CFI and RMSEA are robust values. Samples: 1 = test (Switzerland), 2 = validation (Switzerland), 3 = validation (Australia), 4 = validation (Canada).

behaviour or to maintain discipline in the classroom. Because the content of Items 8 and 9 (control disruptive behaviour and calm disruptive student) and the wording of Items 7 and 8 (prevent disruptive behaviour and control disruptive behaviour) were similar, Items 7 and 9 were retained. Items 11 and 12 were discarded because they had lower loadings and were frequent sources of misfit in previous studies (e.g. Aiello et al., 2018; Alnahdi, 2019; Chao et al., 2016; Malinen, Savolainen, Engelbrecht, et al., 2013; Sharma & Jacobs, 2016).

TABLE 4 CFA and EFA for the TEIP scale and TEIP-SF: factor loadings and reliabilities.

		TEIP scale											
		Step 1: CFA									Step 2: EFA		
		S1			S3			S4			S1		
Nr	Item content	EI	EM	EC	EI	EM	EC	EI	EM	EC	F1	F2	F3
1	Assessments	**0.66**			**0.56**			**0.74**			**0.64**	−0.03	−0.11
2	Explanations	**0.77**			**0.61**			**0.72**			**0.79**	0.00	−0.04
3	Learning tasks	**0.73**			**0.57**			**0.73**			**0.74**	−0.03	0.01
4	Comprehension	**0.69**			**0.52**			**0.75**			**0.51**	−0.16	0.29
5	Capable students	**0.58**			**0.70**			**0.73**			0.38	0.08	0.33
6	Small groups	**0.69**			**0.68**			**0.72**			0.39	0.25	0.30
7	prevent disruption		**0.84**			**0.82**			**0.87**		0.05	**0.75**	0.09
8	Control disruption		**0.89**			**0.89**			**0.91**		−0.02	**0.99**	−0.12
9	Calm student		**0.91**			**0.84**			**0.89**		−0.06	**0.92**	0.00
10	Classroom rules		**0.86**			**0.82**			**0.82**		0.07	**0.75**	0.14
11	Physical aggression		**0.65**			**0.55**			**0.62**		0.05	**0.54**	0.09
12	Clear expectations		**0.71**			**0.67**			**0.71**		0.19	**0.44**	0.31
13	Family assistance			**0.59**			**0.81**			**0.76**	−0.06	0.11	**0.62**
14	Improve learning			**0.65**			**0.85**			**0.75**	0.10	0.06	**0.63**
15	Joint work			**0.80**			**0.63**			**0.72**	0.03	−0.02	0.03
16	Involve parents			**0.79**			**0.72**			**0.77**	0.00	0.03	0.15
17	Collaboration			**0.79**			**0.70**			**0.68**	0.09	0.05	−0.04
18	Laws and policies			**0.57**			**0.59**			**0.62**	0.07	0.08	0.01
	Factor reliability												
	α	0.84	0.91	0.85	0.77	0.90	0.85	0.87	0.91	0.86	0.81	0.91	0.77
	ω	0.84	0.92	0.85	0.78	0.92	0.86	0.87	0.94	0.92	0.82	0.92	0.78

Note: For steps 1 and 3, significant loadings are in bold. For step 2, factor loadings >0.4 are in bold. Additional information: Factor correlations in the correlated-factors models for full-length TEIP, S1: EI/EM = 0.58; EI/EC = 0.76, EM/EC = 0.51; S3: EI/EM = 0.74; EI/EC = 0.81, EM/EC = 0.66; S4: EI/EM = 0.79; EI/EC = 0.85, EM/EC = 0.71. Factor correlations in the correlated-factors models for TEIP-SF, S2: EI/EM = 0.39; EI/EC = 0.63, EM/EC = 0.28; S3: EI/EM = 0.58; EI/EC = 0.81, EM/EC = 0.53; S4: EI/EM = 0.75; EI/EC = 0.75, EM/EC = 0.63.

Abbreviations: EC, efficacy in collaboration; EI, efficacy to use inclusive instructions; EM, efficacy in managing behaviour; S, sample.

For the *Efficacy in Collaboration* factor, Items 15, 16 and 17 were retained. Items 13 and 14 had low factor loadings, and Item 18 did not directly address collaboration. These three items were also frequent sources of misfit in previous studies (e.g. Cardona-Molto et al., 2020; Malinen, Savolainen, Engelbrecht, et al., 2013; Miesera et al., 2019; Mohamed Emam & Al-Mahdy, 2020; Park et al., 2016; Sharma & Jacobs, 2016) so they too were discarded.

Items in the final TEIP-SF are shown in Table 5.

Step 3: The TEIP-SF's reliability and validity

To independently confirm the TEIP-SF's factor structure, its CFAs and factor reliabilities were assessed in three additional validation samples of teachers (Sample 2 from Switzerland, Sample 3 from Australia and Sample 4 from Canada). For the correlated-factors model, the fit indices for all three samples indicated excellent model fit. The reliability of the TEIP-SF's factors was good for EM (ω = 0.89; ω = 0.86; ω = 0.89) and EC (ω = 0.84; ω = 0.80; ω = 0.80) in Samples 2, 3 and 4, respectively, whereas the reliability of EI was questionable in Sample 3 (ω = 0.61), but good in Samples 2 and 4 (ω = 0.81; ω = 0.80).

To determine criterion and convergent validity, we assessed the (sub-)scale score overlap between the TEIP scale and the TEIP-SF, as well as their correlations with the *Beliefs* and *Feelings* towards inclusive education attitude subscales and the overall self-efficacy item. Unweighted (sub-)scale scores in the TEIP scale and the TEIP-SF correlated with *r* values between 0.90 and 0.95 in the three validation samples, indicating a high overlap between the original and short-form scales (Table 6). Also, both the complete and short-form scales displayed similar patterns of correlations between subscale scores. However, the TEIP-SF's subscale correlations were lower in all three samples (especially for the EM subscale),

							TEIP-SF								
							Step 3: CFA								
	S1			S1		S1	S2			S3			S4		
F4	F1	F2	F3	F1	F2	F1	EI	EM	EC	EI	EM	EC	EI	EM	EC
0.21	**0.52**	−0.14	0.29	−0.06	**0.69**	**0.43**	**0.68**			**0.61**			**0.79**		
0.08	**0.69**	−0.08	0.18	0.04	**0.69**	**0.54**	**0.83**			**0.55**			**0.73**		
0.06	**0.69**	−0.09	0.15	0.04	**0.66**	**0.51**	**0.79**			**0.63**			**0.73**		
−0.13	**0.65**	0.23	−0.09	0.37	0.38	**0.65**									
−0.08	**0.54**	0.18	−0.06	0.29	0.34	**0.55**									
−0.05	**0.52**	0.34	−0.02	**0.45**	0.36	**0.70**									
0.03	0.09	**0.78**	0.04	**0.82**	0.05	**0.81**		**0.78**			**0.82**			**0.85**	
0.00	−0.09	**0.93**	0.01	**0.94**	−0.12	**0.77**									
0.05	−0.08	**0.94**	0.06	**0.93**	−0.06	**0.81**		**0.91**			**0.81**			**0.87**	
−0.04	0.12	**0.81**	−0.02	**0.86**	0.02	**0.82**		**0.88**			**0.82**			**0.85**	
0.07	0.07	**0.58**	0.09	**0.60**	0.10	**0.64**									
−0.05	0.33	**0.55**	−0.04	**0.64**	0.18	**0.75**									
0.23	0.23	0.37	0.21	**0.40**	0.36	**0.66**									
0.18	0.38	0.32	0.19	0.38	**0.45**	**0.70**									
0.83	−0.01	−0.01	**0.87**	−0.08	**0.80**	**0.48**			**0.82**			**0.70**			**0.82**
0.72	0.04	0.09	**0.73**	0.03	**0.73**	**0.53**			**0.75**			**0.74**			**0.68**
0.78	0.02	0.03	**0.81**	−0.04	**0.78**	**0.50**			**0.82**			**0.82**			**0.76**
0.48	0.05	0.08	**0.50**	0.04	**0.53**	**0.41**									
0.84	0.84	0.91	0.84	0.91	0.88	0.92	0.80	0.88	0.84	0.62	0.86	0.79	0.78	0.89	0.78
0.85	0.84	0.92	0.85	0.92	0.88	0.93	0.81	0.89	0.84	0.63	0.86	0.80	0.80	0.89	0.80

indicating clearer distinction between domain-specific factors. Moreover, patterns of correlations with attitude factors and the overall self-efficacy item on teaching students with special educational needs were highly comparable between the original and the short-form (sub-) scales. The TEIP-SF even tended to yield higher correlations with the *Beliefs* and *Feelings* attitude scales in Sample 2, while subscale scores for EM and EC (but not EI) correlated lower with the overall self-efficacy item than did the original TEIP scale. However, differences in magnitude were generally negligible (mean difference = 0.03, max. difference = 0.11).

Finally, to determine measurement invariance of the TEIP-SF across the three different validation samples, configural, metric and scalar invariance models were compared (Table 7). Full metric, but only partial scalar measurement invariance could be demonstrated. For partial scalar invariance, the equality constraints regarding the intercepts of two items (Item 1, subscale EI; and Item 16, subscale EC) had to be released.

DISCUSSION

The TEIP scale is a popular scale for assessing TSE about using inclusive practices across three related domains: using inclusive instructions, managing behaviour and collaboration. The present study was undertaken to test the psychometric properties of a short form of the TEIP scale that will enable researchers to measure TSE about using inclusive practices efficiently and concisely. We also sought to ensure a robust factor structure and similar levels of criterion and convergent validities as the full-length TEIP scale. Hence, the short form is intended to complement the full-length TEIP scale, which measures TSE about using inclusive practices more comprehensively.

TABLE 5 Teacher Efficacy for Inclusive Practices scale—Short Form (TEIP-SF).

Factor	Item	Description
Instruction	1	I can use a variety of assessment strategies (e.g. portfolio assessment, modified tests and performance-based assessment)
	2	I am able to provide an alternate explanation or example when students are confused
	3	I am confident in designing learning tasks so that the individual needs of students with disabilities are accommodated
Managing behaviour	7	I am confident in my ability to prevent disruptive behaviour in the classroom before it occurs
	9	I am able to calm a student who is disruptive or noisy
	10	I am able to get children to follow classroom rules
Collaboration	15	I am able to work jointly with other professionals and staff (e.g. aides, other teachers) to teach students with disabilities in the classroom
	16	I am confident in my ability to get parents involved in school activities of their children with disabilities
	17	I can collaborate with other professionals (e.g. itinerant teachers or speech pathologists) in designing educational plans for students with disabilities

Note: The full-length original TEIP scale can be found in Sharma et al. (2012).

The TEIP scale's factor structure

We began by assessing the TEIP scale's factor structure. The original factor solution, with three correlated factors, as proposed by Sharma et al. (2012), has been shown to be very robust across different studies and teacher samples. Slight adaptations, such as discarding certain items, were nevertheless often necessary to reach an adequate model fit (e.g. Aiello et al., 2018; Cardona-Molto et al., 2020; Malinen, Savolainen, & Xu, 2013; Miesera et al., 2019). Initial CFAs using data from our sample of teachers in Switzerland also revealed that the original solution with the three correlated factors as proposed by Sharma et al. (2012) failed to reach an adequate fit to the data. However, the model's fit to our Australian and

TABLE 6 TEIP scale and TEIP-SF total and subscale score correlations.

	TEIP-SF	EI–SF	EM–SF	EC–SF	Beliefs	Feelings	SE SEN
TEIP scale	**0.94/0.95/0.97**	0.85/0.84/0.91	0.80/0.85/0.89	0.84/0.88/0.89	0.34/0.23/0.38	0.45/0.41/0.41	0.44/0.50/0.35
EI	0.83/0.82/0.87	**0.90/0.91/0.94**	0.52/0.57/0.72	0.61/0.63/0.74	0.28/0.26/0.36	0.39/0.43/0.40	0.32/0.42/0.30
EM	0.66/0.76/0.86	0.35/0.42/0.65	**0.95/0.94/0.96**	0.46/0.61/6.4	0.13/0.06/0.29	0.22/0.19/0.29	0.27/0.38/0.30
EC	0.83/0.84/0.84	0.55/0.58/6.1	0.28/0.43/0.55	**0.91/0.91/0.93**	0.42/0.28/0.38	0.51/0.45/0.42	0.49/0.49/0.34
Beliefs	0.40/0.20/0.39	0.34/0.27/0.38	0.13/-0.03/0.31	0.43/0.25/0.32	–	0.71/0.64/0.76	0.33/0.09/0.18
Feelings	0.52/0.41/0.42	0.46/0.43/0.41	0.20/0.13/0.29	0.51/0.43/0.39		–	0.39/0.26/0.25
SE SEN	0.44/0.51/0.33	0.33/0.45/0.34	0.24/0.31/0.29	0.44/0.48/0.23			–

Note: The table displays Pearson correlations. *Diagonal*: Correlations of the original TEIP scale total/subscale scores with TEIP-SF total/subscale scores (displayed in bold). *Upper-right triangle*: Correlations of original TEIP scale total/subscale scores with each other and attitude/overall self-efficacy variables. *Lower-left triangle*: Correlations of TEIP-SF total/subscale scores with each other and attitude/overall self-efficacy variables. The first coefficient is based on validation Sample 2 (Switzerland), the second coefficient is based on validation Sample 3 (Australia) and the third coefficient is based on validation Sample 4 (Canada). Beliefs/Feelings: Attitude toward Inclusion (AIS) subscales.

Abbreviations: EC, efficacy in collaboration; EI, efficacy to use inclusive instructions; EM, efficacy in managing behaviour; SE SEN, self-efficacy in teaching students with special educational needs; SF, short form; TEIP, teacher efficacy for inclusive practices.

TABLE 7 Measurement invariance of the TEIP-SF.

Fit indices	χ^2 (df)	CFI (ΔCFI)	RMSEA (ΔRMSEA)	SRMR (ΔSRMR)	Decision
Configural	107.4 (72)	0.989	0.045	0.036	–
Metric	119.8 (84)	0.988 (−0.001)	0.043 (−0.002)	0.046 (0.010)	Accepted
Scalar	209.6 (96)	0.962 **(−0.026)**	0.071 **(0.028)**	0.064 **(0.018)**	Not accepted
Partial Scalar	146.2 (92)	0.982 (−0.006)	0.051 (0.008)	0.051 (0.005)	Accepted

Note: $N = 911$; Swiss teachers, sample 2: $n = 486$; Australian teachers, sample 3: $n = 149$; Canadian teachers, sample 4: $n = 276$. Bold: critical Δ threshold exceeded. For the partial scalar model, equality constraints of intercepts of Items 1 and 16 were freed.

Canadian samples was acceptable, requiring no modifications. Further inspection of the factor structure using an EFA of the Swiss teacher sample replicated the TEIP scale's proposed three-factor structure, as most of the items loaded on their intended factors. Only Items 13 ('I can assist families in helping their children do well in school') and 14 ('I can improve the learning of a student who is failing') did not fit well—items intended to load on the *Efficacy in Collaboration* factor. Findings from previous studies suggested that this misfit was not due to sample characteristics or item translation issues because these two items frequently caused misfit in a range of different teacher samples from different countries (Cardona-Molto et al., 2020; Malinen et al., 2012; Miesera et al., 2019; Mohamed Emam & Al-Mahdy, 2020; Park et al., 2016). In our Swiss sample, the parallel analysis and the slightly better-fitting four-factor EFA suggested that the TEIP scale captured an additional factor with these two items. Indeed, they originally came from the TSES scale (Tschannen-Moran & Hoy, 2001), where they reflected the *Efficacy in Student Engagement* factor. This might explain the tendency for these two items to cause model–data misfit because they focus on improvements in student learning and not primarily on collaboration. In the TSES, the *Efficacy in Student Engagement* factor is represented by additional items such as 'I can get through to the most difficult students' and 'I can get students to believe that they can do well in schoolwork.' These aspects are also relevant for teaching in inclusive classrooms, and they could be incorporated into future alternative or extended versions of the TEIP scale.

All the TEIP scale's other items fitted our Swiss sample well, with loadings generally >0.5 on their intended factors, with minimal cross-loadings on other factors. Exceptions were Item 6 ('I am confident in my ability to get students to work together in pairs or in small groups') and Item 12 ('I can make my expectations clear about student behavior'), which had cross-loadings >0.3 on other factors. Of these two items, only Item 12 has been observed to cause misfit in several other studies (Alnahdi, 2019; Cardona-Molto et al., 2020; Malinen, Savolainen, Engelbrecht, et al., 2013; Malinen, Savolainen, & Xu, 2013; Sharma & Jacobs, 2016).

These observations were used to develop a short form of the TEIP scale based on the originally proposed three-factor structure. The newly specified TEIP-SF discarded ambivalent items and uses only three items for each factor, with the aim of representing those three factors very efficiently.

The TEIP-SF's efficiency

The TEIP-SF is a good representation of the intended structure of a three-correlated-factors model. This was evidenced by good model fits with three independent validation samples of teachers from Switzerland, Australia and Canada. Because the items now address the factors' intended content more explicitly, those factors are measured more concisely (but also more narrowly) and more distinctly, especially the *EM* factor. Furthermore, good reliability (with $\omega \geq 0.8$) was achieved for all three factors across all three samples, except for the *EI* factor in the Australian teacher sample ($\omega = 0.63$). The TEIP-SF showed a minimal loss of precision compared to the original TEIP scale, as indicated by the short form's high total and subscale score correlations (all $r \geq 0.9$). The TEIP-SF's correlations with criterion validity (self-efficacy about teaching students with disabilities) and convergent validity (AIS *Feelings* and *Beliefs*) variables were similar to those of the original TEIP scale in pattern and magnitude, especially for total scale correlations. Also of great importance is the comparability of the TEIP-SF across different international samples. In contrast to the original TEIP, where poor model fit for the Swiss sample indicated configural noninvariance, the TEIP-SF demonstrated full metric and partial scalar measurement invariance across the three different international teacher samples, allowing for meaningful cross-group comparisons (Byrne et al., 1989; Steenkamp & Baumgartner, 1998).

Limitations

The newly developed TEIP-SF was found to exhibit adequate fit and measurement invariance across three samples of teachers working in different countries. There is no assurance, however, of this new scale's generalizability to other countries or languages. Depending on differences in culture, history, legislation, language or other contextual aspects, the meanings of items (and factors) in the TEIP-SF may change, reflecting similar issues with the original TEIP scale. With only three items per

factor, the TEIP-SF is time efficient, but users should be aware that factor reliability could be an issue (as demonstrated by the *EI* factor's rather low reliability in the Australian sample). If factor reliability is crucial, using the full-length TEIP scale might be a better choice. Also, although the TEIP-SF retains the concepts in the original TEIP scale and covers the most important aspects in TSE about teaching in inclusive classrooms, it may not fully capture a few other relevant aspects. For example, new items could be added that examine student engagement (an aspect covered in the TSES by Tschannen-Moran & Hoy, 2001) or TSE about the creation of socially inclusive classrooms (de Boer et al., 2012). Finally, the TEIP-SF demonstrated similar patterns of criterion and convergent validity to the original full-length TEIP across three different teacher samples. However, because of the sampling process, the three teacher samples are not necessarily representative of the respective national teacher population. Therefore, the generalizability of the results to the respective national contexts cannot be guaranteed either, and further systematic research on the validity, reliability and efficiency of the TEIP-SF is to be recommended.

CONCLUSIONS

To conclude, we posit that the TEIP-SF is a sufficiently valid and reliable scale to measure teacher efficacy in inclusive classrooms, is more time efficient than the original full-length scale and demonstrates a stable factor structure. This could be especially beneficial for (1) studies that face concerns about scale length because they deal with complex research questions and use multiple constructs and (2) studies aiming to compare different samples, where a robust factor structure across samples and the demonstration of measurement invariance are crucial.

FUNDING INFORMATION

This research did not receive any specific grants from funding agencies in the public, commercial or not-for-profit sectors.

ACKNOWLEDGMENTS

Open access funding provided by Padagogische Hochschule Bern.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflicts of interest.

DATA AVAILABILITY STATEMENT

Data of the Swiss teacher sample are available at the SWISSUbase repository: Sahli Lozano, Caroline, Wüthrich, Sergej, Baumli, Nadja & Büchi, Jonas (2023). Daten von Berner und Deutschschweizer Lehrpersonen und sonderpädagogischen Fachpersonen zu Einstellungen, Bedenken und Selbstwirksamkeit gegenüber schulischer Inklusion (ISASI) (2019–2022) (1.0.0) [Datensatz]. FORS Datenservice. 10.48573/cbjp-wk67. Data from Australian and Canadian teachers are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

Ethics approval was not required.

ORCID

C. Sahli Lozano https://orcid.org/0000-0002-3957-8745
S. Wüthrich https://orcid.org/0000-0002-3301-8587
U. Sharma https://orcid.org/0000-0002-5198-9379
T. Loreman https://orcid.org/0000-0002-6666-1720
C. Forlin https://orcid.org/0000-0002-6438-9591

REFERENCES

Aiello, P., Pace, E.M., Dimitrov, D.M. & Sibilio, M. (2018) A study on the perceptions and efficacy towards inclusive practices of teacher trainees. *Italian Journal of Educational Research*, 19, 13–28.

Alnahdi, G.H. (2019) The Arabic version of the teacher efficacy for inclusive practices (TEIP-AR) scale: a construct validity study. *Cogent Education*, 6(1), 1–11. Available from: https://doi.org/10.1080/2331186X.2019.1618516

Bandura, A. (1977) Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. Available from: https://doi.org/10.1037/0033-295X.84.2.191

Byrne, B.M., Shavelson, R.J. & Muthén, B. (1989) Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466. Available from: https://doi.org/10.1037/0033-2909.105.3.456

Cardona-Molto, M.C., Ticha, R. & Abery, B.H. (2020) The Spanish version of the teacher efficacy for inclusive practice (TEIP) scale: adaptation and psychometric properties. *European Journal of Educational Research*, 9(2), 809–823. Available from: https://doi.org/10.12973/eu-jer.9.2.809

Chao, C.N.G., Forlin, C. & Ho, F.C. (2016) Improving teaching self-efficacy for teachers in inclusive classrooms in Hong Kong. *International Journal of Inclusive Education*, 20(11), 1142–1154. Available from: https://doi.org/10.1080/13603116.2016.1155663

Chen, F.F. (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. Available from: https://doi.org/10.1080/10705510701301834

Cieciuch, J. & Davidov, E. (2015) Establishing measurement invariance across online and offline samples. A tutorial with the software packages Amos and Mplus. *Studia Psychologica*, 14, 83. Available from: https://doi.org/10.5167/UZH-170024

de Boer, A., Pijl, S.J. & Minnaert, A. (2012) Students' attitudes towards peers with disabilities: a review of the literature. *International Journal of Disability, Development and Education*, 59(4), 379–392. Available from: https://doi.org/10.1080/1034912X.2012.723944

Dimitrov, D.M. (2012) *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA: American Counseling Association.

Gebhardt, M., Miesera, S., Weidenhiller, P. & Jungjohann, J. (2018) *Einstellung, Bedenken und Selbstwirksamkeitserwartung von Lehramtsstudierenden zur schulischen Inklusion*. Dortmund,

Germany: Technische Universität Dortmund. Available at: https://eldorado.tu-dortmund.de/handle/2003/36870 [Accessed 1 December 2020].

Gibson, A.M. & Bowling, N.A. (2020) The effects of questionnaire length and behavioral consequences on careless responding. *European Journal of Psychological Assessment*, 36(2), 410–420. Available from: https://doi.org/10.1027/1015-5759/a000526

Hayes, A.F. & Coutts, J.J. (2020) Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, 14(1), 1–24. Available from: https://doi.org/10.1080/19312458.2020.1718629

Howard, M.C. (2016) A review of exploratory factor analysis decisions and overview of current practices: what we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1), 51–62. Available from: https://doi.org/10.1080/10447318.2015.1087664

Hu, L. & Bentler, P.M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. Available from: https://doi.org/10.1080/10705519909540118

Malinen, O.-P., Savolainen, H., Engelbrecht, P., Xu, J., Nel, M., Nel, N. et al. (2013) Exploring teacher self-efficacy for inclusive practices in three diverse countries. *Teaching and Teacher Education*, 33, 34–44. Available from: https://doi.org/10.1016/j.tate.2013.02.004

Malinen, O.-P., Savolainen, H. & Xu, J. (2012) Beijing in-service teachers' self-efficacy and attitudes towards inclusive education. *Teaching and Teacher Education*, 28(4), 526–534. Available from: https://doi.org/10.1016/j.tate.2011.12.004

Malinen, O.-P., Savolainen, H. & Xu, J. (2013) Dimensions of teacher self-efficacy for inclusive practices among mainland Chinese pre-service teachers. *Journal of International Special Needs Education*, 16(2), 82–93. Available from: https://doi.org/10.9782/2159-4341-16.2.82

McNeish, D. (2018) Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. Available from: https://doi.org/10.1037/met0000144

Miesera, S., DeVries, J.M., Jungjohann, J. & Gebhardt, M. (2019) Correlation between attitudes, concerns, self-efficacy and teaching intentions in inclusive education evidence from German pre-service teachers using international scales. *Journal of Research in Special Educational Needs*, 19(2), 103–114. Available from: https://doi.org/10.1111/1471-3802.12432

Mohamed Emam, M. & Al-Mahdy, Y.F.H. (2020) Teachers' efficacy for inclusive practices in the Sultanate of Oman: effect of gender and teaching experience. *School Psychology International*, 41(2), 170–192. Available from: https://doi.org/10.1177/0143034319895062

Narkun, Z. & Smogorzewska, J. (2019) Studying self-efficacy among teachers in Poland is important: polish adaptation of the teacher efficacy for inclusive practice (TEIP) scale. *Exceptionality Education International*, 29, 110–126. Available from: https://doi.org/10.5206/eei.v29i2.9405

O'Connor, B.P. (2022) *EFA.Dimensions: exploratory factor analysis functions for assessing dimensionality*. Available at: https://CRAN.R-project.org/package=EFA.dimensions [Accessed 29 September 2022].

Opoku, M.P. (2021) Exploring the intentions of school leaders towards implementing inclusive education in secondary schools in Ghana. *International Journal of Leadership in Education*, 1–21. Available from: https://doi.org/10.1080/13603124.2021.1889034

Osborne, J.W. (2014) *Best practices in exploratory factor analysis*. Louisville, KY: CreateSpace Independent Publishing Platform.

Papenberg, M. & Klau, G.W. (2021) Using anticlustering to partition data sets into equivalent parts. *Psychological Methods*, 26(2), 161–174. Available from: https://doi.org/10.1037/met0000301

Park, M.-H., Dimitrov, D.M., Das, A. & Gichuru, M. (2016) The teacher efficacy for inclusive practices (TEIP) scale: dimensionality and factor structure. *Journal of Research in Special Educational Needs*, 16(1), 2–12. Available from: https://doi.org/10.1111/1471-3802.12047

Pit-ten Cate, I.M., Schwab, S., Hecht, P. & Aiello, P. (2019) Editorial: Teachers' attitudes and self-efficacy beliefs with regard to inclusive education. *Journal of Research in Special Educational Needs*, 19(S1), 3–7. Available from: https://doi.org/10.1111/1471-3802.12480

R Core Team. (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing. Available at: https://www.R-project.org/ [Accessed 31 March 2020].

Revelle, W. (2022) *Psych: procedures for psychological, psychometric, and personality research*. Evanston, IL: Northwestern University. Available at: https://CRAN.R-project.org/package=psych [Accessed 21 May 2020].

Rosseel, Y. (2012) Lavaan: an R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. Available from: https://doi.org/10.18637/jss.v048.i02

Rosseel, Y., Jorgensen, T.D., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L. et al. (2022) *lavaan: Latent Variable Analysis* (0.6-10). Available at: https://CRAN.R-project.org/package=lavaan [Accessed 22 February 2022].

Savolainen, H., Engelbrecht, P., Nel, M. & Malinen, O.-P. (2012) Understanding teachers' attitudes and self-efficacy in inclusive education: implications for pre-service and in-service teacher education. *European Journal of Special Needs Education*, 27(1), 51–68. Available from: https://doi.org/10.1080/08856257.2011.613603

Schmitt, T.A., Sass, D.A., Chappelle, W. & Thompson, W. (2018) Selecting the "best" factor structure and moving measurement validation forward: an illustration. *Journal of Personality Assessment*, 100(4), 345–362. Available from: https://doi.org/10.1080/00223891.2018.1449116

Sharma, U. & Jacobs, D.K. (2016) Predicting in-service educators' intentions to teach in inclusive classrooms in India and Australia. *Teaching and Teacher Education*, 55, 13–23. Available from: https://doi.org/10.1016/j.tate.2015.12.004

Sharma, U., Loreman, T. & Forlin, C. (2012) Measuring teacher efficacy to implement inclusive practices. *Journal of Research in Special Educational Needs*, 12(1), 12–21. Available from: https://doi.org/10.1111/j.1471-3802.2011.01200.x

Skaalvik, E.M. & Skaalvik, S. (2007) Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, 99(3), 611–625. Available from: https://doi.org/10.1037/0022-0663.99.3.611

Steenkamp, J.E.M. & Baumgartner, H. (1998) Assessing measurement invariance in cross-National Consumer Research. *Journal of Consumer Research*, 25(1), 78–107. Available from: https://doi.org/10.1086/209528

Swiss Federal Statistical Office. (2018) *Switzerland's population 2017* (No. 348-1700). Available at: https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/publikationen.assetdetail.6606499.html [Accessed 4 April 2019].

Tanrıverdi, A. & Özokçu, O. (2018) The psychometric properties of the Turkish version of the teacher efficacy for inclusive practices (TEIP) scale. *Educational Research and Reviews*, 13(18), 654–663. Available from: https://doi.org/10.5897/ERR2018.3504

Tschannen-Moran, M. & Hoy, A.W. (2001) Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783–805. Available from: https://doi.org/10.1016/S0742-051X(01)00036-1

Vogiatzi, C.-A., Charitaki, G., Kourkoutas, E. & Forlin, C. (2022) The teacher efficacy for inclusive practices (TEIP) scale: further evidence for construct validity in Greek-speaking teachers. *Prospects*, 52, 387–403. Available from: https://doi.org/10.1007/s11125-022-09605-w

Watkins, M.W. (2018) Exploratory factor analysis: a guide to best practice. *Journal of Black Psychology*, 44(3), 219–246. Available from: https://doi.org/10.1177/0095798418771807

Yada, A., Björn, P.M., Savolainen, P., Kyttälä, M., Aro, M. & Savolainen, H. (2021) Pre-service teachers' self-efficacy in implementing inclusive practices and resilience in Finland. *Teaching and Teacher Education*, 105, 103398. Available from: https://doi.org/10.1016/j.tate.2021.103398

Yada, A., Leskinen, M., Savolainen, H. & Schwab, S. (2022) Meta-analysis of the relationship between teachers' self-efficacy and attitudes toward inclusive education. *Teaching and Teacher Education*, 109, 103521. Available from: https://doi.org/10.1016/j.tate.2021.103521

Yada, A., Tolvanen, A., Malinen, O.-P., Imai-Matsumura, K., Shimada, H., Koike, R. et al. (2019) Teachers' self-efficacy and the sources of efficacy: a cross-cultural investigation in Japan and Finland. *Teaching and Teacher Education*, 81, 13–24. Available from: https://doi.org/10.1016/j.tate.2019.01.014

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. Available from: https://doi.org/10.3102/0034654315626801

How to cite this article: Sahli Lozano, C., Wüthrich, S., Baumli, N., Sharma, U., Loreman, T. & Forlin, C. (2023) Development and validation of a short form of the Teacher Efficacy for Inclusive Practices Scale (TEIP-SF). *Journal of Research in Special Educational Needs*, 00, 1–14. Available from: https://doi.org/10.1111/1471-3802.12607