



## RESEARCH ARTICLE

# Validating the use of the 24-item long version and the 12-item short version of the *Teachers' Sense of Efficacy Scale* (TSES) for measuring teachers' self-efficacy in Macao (SAR) for inclusive education [version 1; peer review: 1 approved with reservations, 1 not approved]

Elisa Monteiro <sup>1</sup>, Chris Forlin<sup>2</sup>

<sup>1</sup>School of Education, University of Saint Joseph, Macao, Macao, Macao

<sup>2</sup>International Inclusive Education Consultant, Bayswater, WA, Australia

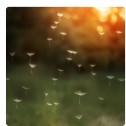
**V1** **First published:** 09 Jun 2020, 2:36  
<https://doi.org/10.35241/emeraldopenres.13541.1>  
**Latest published:** 09 Jun 2020, 2:36  
<https://doi.org/10.35241/emeraldopenres.13541.1>

## Abstract

Validation of the *Teachers' Sense of Efficacy Scale* (TSES) for use with teachers in Macao (SAR) was undertaken to determine its usefulness as a measure of teacher self-efficacy for inclusive education. This paper discusses the results found by analyzing various versions of the TSES and TSES-C in a Chinese format with 200 pre-service teachers in Macao (SAR). Psychometric analyses were undertaken to investigate the validity of the existing scales and the three and two factor solutions. The results indicated a preferred 9-item version that produced improved factor loadings and reliabilities. The use of a relatively quick and short scale to measure such a complex phenomenon as teacher self-efficacy is discussed. Issues are raised regarding generalizability of scales and the impact of culture, demographics, and edifying issues that may impact on the usefulness of such scales.

## Keywords



Inclusive education, Macao, self-efficacy, scale development




This article is included in the [Quality Education for All](#) gateway.

## Open Peer Review

Approval Status  

	1	2
<b>version 1</b> 09 Jun 2020	 <a href="#">view</a>	 <a href="#">view</a>

1. **Nikolaos Tsigilis**, Aristotle University of Thessaloniki, Thessaloniki, Greece
2. **Stewart Martin** , University of Hull, Kingston upon Hull, United Kingdom  
University of Saint Joseph, Macau SAR, Peoples Republic Of China

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Elisa Monteiro ([elisa@usj.edu.mo](mailto:elisa@usj.edu.mo))

**Author roles:** **Monteiro E:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Forlin C:** Formal Analysis, Methodology, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by the Macao Foundation [MF/2018/46].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2020 Monteiro E and Forlin C. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Monteiro E and Forlin C. **Validating the use of the 24-item long version and the 12-item short version of the Teachers' Sense of Efficacy Scale (TSES) for measuring teachers' self-efficacy in Macao (SAR) for inclusive education [version 1; peer review: 1 approved with reservations, 1 not approved]** Emerald Open Research 2020, 2:36 <https://doi.org/10.35241/emeraldopenres.13541.1>

**First published:** 09 Jun 2020, 2:36 <https://doi.org/10.35241/emeraldopenres.13541.1>

## Introduction

A consultation document to make provisions to Macao's Decree-Law on special education, to promote a more inclusive school system, was launched by the Macao Government in 2015 (Direcção dos Serviços de Educação e Juventude (DSEJ, 2015)). Since the launch, several initiatives have taken place to lead schools toward more fully inclusive practices (Teixeira *et al.*, 2018). The Macao educational system has undergone a series of changes, not only with education reforms, but also in development of policies for inclusive education. These developments can be seen in initiatives such as the *Ten-Year Plan* (2011–2020) for development of non-tertiary education (DSEJ, 2014). The initiatives target attention to improve student learning through a student-centred approach, strengthen students' critical thinking skills and long-life skills, as well as improve teachers' working conditions and professional development.

As a result of education reform, many schools have established more comprehensive mechanisms for education quality assurance and school effectiveness, and inclusive education, although this is still in its infancy.

In spite of the provision of the law on inclusive education, achieving inclusion presents a myriad of challenges for schools and teachers in Macao. Among some of these challenges are attitudes and cultural beliefs toward differences and diversity (Chao *et al.*, 2016; Correia *et al.*, 2019); placements of students with special education needs (Teixeira *et al.*, 2018), teacher support and inadequate preparation of teachers (Monteiro *et al.*, 2018). To ensure equal opportunity to education for all learners including those with diverse abilities, disabilities and special needs, requires teachers who believe they are capable of supporting them. By developing an effective and valid measure of self-efficacy it will be possible to measure teachers' perceived readiness for inclusive education over time; to safeguard that they are ready to become inclusive practitioners.

### Improving teacher preparation for inclusion

Commencing in 2012, legislation regarding teacher recruitment and teacher qualification has been implemented with implications for teacher education and appropriate professional development (DSEJ, 2014). One outcome of these reviews was to improve the professional learning of teachers so that they possess an adequate level of knowledge and teaching skills to teach students with diverse learning needs and to meet the demands of the curriculum reform. The new qualification and certification requirements mandated in response to changes in policies and in educational practice, as well as changes in teacher education standards, initiated several actions in teacher education programs to prepare teachers to facilitate the curriculum reform and to better prepare them for inclusive education. These required changes led directly to a major private university in Macao revising its teacher education program to meet these demands (Monteiro & Forlin, *in press*). The one-year Post-Graduate Diploma in Education (PGDE) was reconceptualized to ensure that educational inclusion was addressed within its core courses based on the acknowledgement of inclusion and the importance of enhanced field experiences.

The adequacy of changing training for teachers, however, does not always result in immediately improved attitudes towards inclusion (Chong *et al.*, 2007; Forlin *et al.*, 2014; Forlin & Chambers, 2010). According to Singh *et al.*, (2019), there are four major types of teacher education. These include those related to teachers who are reflective, effective, enquiring and transformative. Determining an ideal approach will depend heavily on the context, demographics, and Government expectations within any given region. As the study by Monteiro and Forlin suggests, nevertheless, an effective curriculum for teacher education should be an important resource to support changes in Macao's education system.

### Teacher self-efficacy

Rooted in Bandura's (1977) social cognitive theory, teacher self-efficacy has proven to play an essential role in improving teachers' pedagogical practices. Results of many empirical studies have been consistent with Bandura's theory that teachers' efficacy beliefs are related to the effort teachers put into their teaching. Studies have shown teacher efficacy to be strongly related to a wide range of instructional qualities (Holzberger *et al.*, 2013; Künsting *et al.*, 2016), and student academic achievement (Guo *et al.*, 2010). Bandura's theory of self-efficacy suggests efficacy beliefs are influenced by mastery experience, physiological and emotional states, social persuasion, vicarious experiences, and cognitive factors, and are most malleable early in learning (Bandura, 1977). Several studies have also looked at teacher-efficacy during pre-service teacher education and have reported increases during training. For example, a study by Main & Hammond (2008) reported higher self-efficacy beliefs among prospective teachers in behavior management after the practicum.

In the construct of teacher self-efficacy in inclusive schools it has been found that different cultural contexts have indicated that self-efficacy in teaching is a multidimensional construct (e.g. Loreman *et al.*, 2013; Monteiro *et al.*, 2018). In particular, attitudes towards inclusive education are not constant but are influenced continually and significantly by the social milieu (Lüke & Grosche, 2018). Tschannen-Moran & Hoy (2006) similarly reported that teacher self-efficacy is context-specific with contextual factors, such as teaching resources and interpersonal support, found to be more prominent in the perceived self-efficacy of pre-service teachers. Teacher self-efficacy has shown to be a salient variable in predicting teachers' capability to execute effective practices and transformation within educational reforms.

Previously in Macao, teachers have been found to disagree with the core values of inclusive education (HKIEd, 2012). This view continued to be held in 2015 when they similarly lacked positive views about inclusive education (Cheung *et al.*, 2015). Within the quite dramatic educational changes occurring in Macao, teacher self-efficacy is going to be critical for ensuring effective engagement in inclusive education and ongoing sustainability. For teachers to better respond to the educational reforms in Macao, teacher education and training must equip novice teachers with the necessary pedagogical skills and

content knowledge to ensure that teachers are prepared and willing to support students with disabilities in general education classrooms (Monteiro *et al.*, 2018). Together with relevant skills and knowledge, teachers need efficacy for teachers to be confident in their ability to enact effective teaching practices. The beliefs of teachers are critical as they have been found to be concomitant with their inclination to endorse inclusive education (Copfer & Specht, 2014) and to their professed self-efficacy (Chao *et al.*, 2016).

### Measuring teacher self-efficacy

The measurement of self-efficacy has been conceptualized in many ways. Some of these include the Rand measure, grounded in Rotter's social learning theory (Rotter, 1966). The first studies of efficacy, conducted by the Rand researchers, included a questionnaire with two efficacy items. Teachers' level of efficacy was determined by calculating a total score for responses to two items: (1) "When it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment," and (2) "If I try really hard, I can get through to even the most difficult or unmotivated students" (Armor *et al.*, 1976). Shortly after this study was published, another study by Rand researchers was conducted to examine federal funded initiatives designed to spread innovative practices in public schools. This found teacher efficacy to be strongly related to project goals achieved, the amount of teacher change, and continuation of methods and materials (Berman *et al.*, 1977). The results of the two Rand studies sparked interest in the construct of teacher efficacy; however, other researchers found problems with the reliability of the two-item scale. To improve reliability and validity, other researchers, such as Gibson & Dembo (1984) and Rose & Medway (1981) developed more comprehensive measures of teacher efficacy. Other measures include the Webb scales (Ashton *et al.*, 1982) that sought to expand reliability of the Rand measure. Whilst these measures were grounded in Rotter's social learning theory, another self-efficacy conceptual strand manifested based on Bandura's social cognitive theory (Bandura, 1977) and his construct of self-efficacy. Bandura (1977) proposed a model derived from four sources of information: performance accomplishments, vicarious experience, verbal persuasion, and physiological states (p. 191).

Since the development of the Rand measure and other instruments, researchers have engaged in evolving increasing valid and reliable measures of teacher sense of efficacy and several measures have been established over the years. Tschannen-Moran *et al.* (1998) examined the theoretical and empirical underpinnings of teacher efficacy as a construct and psychometric properties of existing instruments to bring coherence to the construct and its measurement. Their model suggested that a valid measure of teacher efficacy must assess personal competence and an analysis of the task (e.g. resources and constraints in teaching contexts) and thus they developed the *Teacher's Sense of Efficacy Scale* (TSES) from three separate studies (Tschannen-Moran & Hoy, 2001). The TSES consists of a long (24-item) and short (12-item) scale that represent three related subscales: efficacy for instructional strategies, efficacy for classroom management, and efficacy for student engagement.

The TSES long and short versions have become the predominant measure of teacher efficacy with practicing and pre-service teachers in many regions. These measures are supported by empirical evidence reported in several studies conducted in the United States (Duffin *et al.*, 2012; Klassen *et al.*, 2009), Canada (Klassen *et al.*, 2009), Singapore (Klassen *et al.*, 2009), Greece (Tsigilis *et al.*, 2010), and Korea (Klassen *et al.*, 2009) among others, providing strong score validity evidence to support the factors (Klassen *et al.*, 2009). The scale has been translated into several languages. The Chinese 12-item short form version (TSES-C) has been used in Hong Kong and Macao, although several studies have reported different and varied factor structures (Cheung, 2006; Hui *et al.*, 2006; Kennedy & Hui, 2004).

While pre-service preparation for teaching has changed in Macao to incorporate inclusive education (Monteiro & Forlin, *in press*), as yet there is no valid and reliable measure for determining whether this is meeting the needs of improving teacher self-efficacy for inclusive practice. The purpose of this study was to validate the TSES in both the long and short versions for use with teachers in Macao to determine its usefulness as a measure of teacher self-efficacy for inclusive education. This study examined various versions of the TSES and TSES-C in a Chinese format with 200 pre-service teachers in Macao (SAR).

### Method

#### The Teachers' Sense of Efficacy (TSES) scale

The original *Teachers' Sense of Efficacy* (TSES) scale was developed in two English formats: a 24-item long form and a 12-item short form (Tschannen-Moran *et al.*, 1998). The scales were designed to measure the degree to which a teacher perceives they can affect student performance. Validation of both scales identified three factors pertaining to efficacy in student engagement, instructional strategies, and classroom management (Tschannen-Moran & Woolfolk Hoy, 2001). Reliabilities above .81 were recorded for the full scale and three factors for both the long and short versions.

While the original 24 and 12-item versions of the scales were in English, further research considered a Chinese form of the 12-item short version of this scale (TSE-C; Cheung, 2006; Kennedy & Hui, 2004). The original Chinese version was designed and validated for use by in-service teachers in Hong Kong (Kennedy & Hui, 2004). This version was subsequently validated for use by pre-service teachers in Macao (Cheung, 2006b). Kennedy & Hui (2004) extracted the 12-items into two scales focusing on efficacy in learning and teaching, and efficacy in classroom management. These were confirmed in the Cheung (2006) study, although it was reported that Item 11 *How much can you assist families in helping their children do well in school?* had the lowest mean and was excluded during the confirmatory factor analysis (CFA).

This research aimed to compare findings using the short 12-item version (TSE-C) that had been previously validated for use in Chinese, with the original 24-item scale in a newly translated Chinese version. Both versions underwent extensive validation procedures before developing a 9-item short two

factor version that had improved factor loadings and reliabilities. These were used to review and compare perceptions of efficacy from pre-service teachers after completing a one-year PGDE program in Macao. Translation of the 24-item version was achieved through back translation of the original instrument from English to Chinese and then from Chinese to English.

### Participants

The study took place in one institution in Macao between December 2018 and August 2019. The participants were all pre-service teachers undertaking a one-year teacher preparation program (PGDE) at a private university in Macao. At the administrative stage all participants had completed the structured 90 hours of teaching practice and were at the end of the program offered in the academic year 2018/2019. In total, 36 of the group were men and 84 women (total, 120); 50 were below 25 years of age, and 70 were 26 years or above. The pre-service teachers were preparing to teach at different levels including 44 in kindergarten, 58 in primary and 18 in secondary schools.

In June 2019, during the last course of the program, all pre-service teachers undertaking the program were invited to participate in the study. The participants were asked to complete hard copies of the long questionnaire at the beginning of a class and submission of the questionnaire confirmed agreement to participate. In total, 120 questionnaires were collected. Not all results utilize the 120 data set due to some missing data.

*Sample size.* Conventionally, the minimum number of participants and subject-to-item ratio are two conventional approaches for handling the issue of sample size for exploratory factor analysis (EFA). Two commonly cited reviews (Osborne, 2014; Osborne *et al.*, 2008), though, have pointed out that subject-to-item ratio is more preferable in practice. The majority of published studies adopt the subject-to-item ratio of 10:1. Accordingly, the present study employs the subject-to-item ratio of 10:1.

### S

Internal consistency, exploratory factor analysis (EFA) and confirmatory factor analysis were performed. Data cleansing and assumption testing (e.g. normal distribution, outliers) were conducted before the data analysis. No test-retest reliability and inter-rater reliability was planned or calculated. The participants were asked to answer the Chinese version of the long questionnaire.

The short version of the TSES consisted of only 12 items from the long version (i.e. *Efficacy in Student Engagement*: Items 4, 9, 6, 22; *Efficacy in Instructional Strategies*: 11, 18, 20, 23; and *Efficacy in Classroom Management*: items 3, 15, 13, 16). These 12 items were analyzed accordingly.

IBM SPSS (Version 24.0) and IBM Amos (Version 24.0) was used for analysis.

### Ethical considerations

The study was approved by the university. All participants were informed of the purpose of the research, procedures,

confidentiality, and how the researcher would maintain the privacy of participants through the use of anonymized data. Participants provided informed consent to participate in the research project, and submission of the questionnaire confirmed their participation in the study. The research conformed to the ethical requirements of the university.

## Results

### Long version of the TSES scale

The translated Chinese version of the 24-item scale had very strong reliability at .93 from 118 returned questionnaires. The three existing factors (n=8 items in each) also had high reliability with Cronbach's alpha indicating *efficacy in student engagement* .81; *efficacy in classroom management* .83; and *efficacy in instructional strategies* .88. The long version of the Chinese scale contained good reliabilities, therefore, validating its use in a Chinese format. It is noteworthy that the sample size of the present study is inadequate for factor analysis since it did not fulfill the subject-to-item ratio of 10:1 (Wolf *et al.*, 2013).

Before data analysis, assumption of normality was not violated, and outliers were not found. Descriptive statistics including the means and standard deviations of the three factors were obtained. A one-sample t-test was conducted to compare the means of the three factors for significant differences between them. Means of the *efficacy in instructional strategies* and the *efficacy in classroom management* were used as the test values to compare against the mean of *efficacy in student engagement* and against each other. There was no significant difference between the means of *efficacy in student engagement* ( $M = 6.50, sd = 0.83$ ) and of *efficacy in instructional strategies* ( $M = 6.56, sd = 0.80$ ;  $t(119) = -0.78, p = 0.44$ , two-tailed). Non-significant difference between the means of *efficacy in student engagement* and of *efficacy in classroom management* ( $M = 6.47, sd = 0.94$ ;  $t(119) = 0.41, p = 0.68$ , two-tailed) was also observed. Moreover, the mean of *efficacy in instructional strategies* did not differ significantly from the mean of *efficacy in classroom management* ( $M = 6.47, S.D = 0.94$ ;  $t(119) = 1.30, p = 0.20$ , two-tailed) (Table 1).

The 24-item Chinese form and the three factors were subsequently used to identify any participant differences for the three

**Table 1. Mean differences between the three factors for the 24-item Chinese form of the Teachers' Sense of Efficacy Scale (TSES).**

	N	M	SD	Test value	t-value	df	p-value
Efficacy in Student Engagement	120	6.50	0.83	6.56	-0.78	119	0.44
Efficacy in Instructional Strategies	120	6.56	0.80	6.47	0.41	119	0.68
Efficacy in Classroom Management	120	6.47	0.94	6.47	1.30	119	0.20
Overall Teacher Efficacy	120	6.51	0.76				



independent variables of gender (male, female); age (25 years or below; 26–35 years; 36–45 years; 46 years & above); and teaching level (kindergarten, primary, secondary).

**Gender.** A one-way multivariate analysis of variance (MANOVA) was performed to investigate gender differences. Preliminary data scanning showed that the assumptions were not violated. There was a statistically significant difference between men and women on the total scale score,  $F(1, 117) = 3.45, p = 0.02$ , Wilks' Lambda = 0.92; partial eta squared = 0.08. When the results for the three factors were considered separately, the only difference to reach statistical significance using a Bonferroni adjusted alpha level of 0.017, was *efficacy in instructional strategies*,  $F(1, 117) = 7.75, p = 0.006$ , partial eta squared = 0.06. An inspection of the mean scores indicated that men reported higher levels of *efficacy in instructional strategies* ( $M = 6.87, S.D = 0.72$ ) than women ( $M = 6.43, S.D = 0.80$ ) with a moderate effect size of Cohen's  $d = .58$ .

**Age.** A one-way between groups MANOVA was performed to investigate differences across the three factors for the four age groups (25 years or below; 26–35 years; 36–45 years; 46 years & above). Before data analysis, preliminary assumption testing was conducted, and no violations were found. Results

indicated that there were no statistical differences among the four age groups on the three factors,  $F(3, 116) = 0.96, p = 0.81$ , Wilks' Lambda = 0.58; partial eta squared = 0.015. Due to small group sizes, data were subsequently recalculated into two groups of 25 years or below and 26 years and above to observe differences across the three factors for the two redefined age groups. Before data analysis, preliminary assumption testing was conducted, and no violations were found. A series of one-way ANOVAs were performed. Results indicated no statistical significance between the two age groups on the three factors and the total scale score  $F(1, 118) = 0.40, p = 0.75$ , Wilks' Lambda = 0.99; partial eta squared = 0.01.

**Teaching level.** Since the assumption of homogeneity of variance-covariance matrices for one-way MANOVA was violated (i.e. the sig. value was 0.001), a series of one-way ANOVAs were conducted to examine for significant differences across the three factors and the total scale score. Participants were divided into three groups according to their teaching level (Group 1: Kindergarten; Group 2: Primary; Group 3: Secondary). No significant differences were observed (Table 2).

#### Short version of the TSE scale (12 items)

This consists of 3- and 2-factor versions

**Table 2. Efficacy according to teaching level for the three factors and overall total scale of the Teachers' Sense of Efficacy Scale (TSES).**

	N M SD					S.S.	df	M.S.	F	p-value
Efficacy in Student Engagement	Kindergarten	43	6.61	0.78	Between Groups	0.85	2	0.42	0.60	0.55
	Primary	58	6.47	0.87	Within Groups	81.26	116	0.70		
	Secondary	18	6.38	0.85	Total	82.10	118			
	Total	119	6.51	0.83						
Efficacy in Instructional Management	Kindergarten	43	6.42	0.77	Between Groups	2.31	2	1.15	1.88	0.16
	Primary	58	6.62	0.80	Within Groups	71.31	116	0.61		
	Secondary	18	6.83	0.74	Total	73.62	118			
	Total	119	6.58	0.79						
Efficacy in Classroom Management	Kindergarten	43	6.63	0.81	Between Groups	1.44	2	0.72	0.82	0.44
	Primary	58	6.41	1.07	Within Groups	102.12	116	0.88		
	Secondary	18	6.38	0.76	Total	103.57	118			
	Total	119	6.48	0.94						
Overall Teacher Efficacy	Kindergarten	43	6.55	0.75	Between Groups	0.07	2	0.03	0.06	0.94
	Primary	58	6.50	0.82	Within Groups	68.67	116	0.59		
	Secondary	18	6.52	0.63	Total	68.74	118			
	Total	119	6.52	0.76						

The original 3-factor structure by Tschannen-Moran *et al.* (1998). Using the 3-factor structure identified in the original 12-item scale by Tschannen-Moran *et al.* (1998), Cronbach's alpha was calculated and found to be good for the overall scale at .88 and for the factor of *class management* at .82. Cronbach's alpha for the remaining two factors was only moderate at .73 for both *student engagement* and *instructional strategies*. Compared to the long scale, the reliabilities across all factors were lower. Assumption of normality was not violated and outliers were not found. Descriptive statistics including the means and standard deviations of the three factors were, therefore, obtained. Similar to the long version of the scale, a series of one sample t-tests or ANOVAs found no significant differences between the three factors or for the three independent variables.

The 2-factor structure identified by Kennedy & Hui (2004) and validated further by Cheung (2006). The present study attempted to validate the TSE-C by CFA with AMOS 24 (Arbuckle, 2018), whereas previous validations were performed by EFA (Tsui & Kennedy, 2009). The method of estimation used in both EFAs and CFAs was Maximum Likelihood (Byrne, 2010). The data collected were investigated for fit with the models built in the Chinese sample (i.e. Cheung, 2006; Tsui & Kennedy, 2009). Five model fit indices were chosen as the references of the goodness of model fit (Dragan & Toplsek, 2014; Sivan *et al.*, 2014). When three out of five indices met the acceptable cut-off point, the model was considered as marginal or acceptable; less than three out of five, poor; more than three out of five, good.

Initial attempts to fit Cheung's (2006) model failed because of the presence of missing data as indicated. Only one piece of missing data on Item 3 was identified and the Little's MCAR test showed that this was missing completely at random (i.e.  $\chi = 7.38$ ,  $df = 11$ ,  $p = .79$ ). After selecting the *estimate means and intercept* option from AMOS, model fit indices were shown (Byrne, 2010). Poor fit was revealed with goodness-of-fit indexes:  $\chi(53) = 118.87$ ,  $p < .001$ ;  $CMIN/DF = 2.24$ ,  $CFI = .88$ ,  $NFI = .81$ ,  $TLI = .82$  and  $RMSEA = .10$  (90% CI of  $RMSEA = .08-.13$ ). The factor loadings for *classroom management* (Items 1, 6, 7, 8) ranged from .64 to .74 and *learning and teaching* (Items 2, 3, 4, 5, 9, 10, 11, 12), from .46 to .77. The lowest factor loading was of Item 2 ( $r = .46$ ) due to its low inter-item correlation. As shown in Table 3 (Model 7), after deleting Item 2, model fits improved but were still unsatisfactory or at best marginal:  $\chi(43) = 95.91$ ,  $p < .001$ ;  $CMIN/DF = 2.23$ ,  $CFI = .90$ ,  $NFI = .83$ ,  $TLI = .84$  and  $RMSEA = .10$  (90% CI of  $RMSEA = .07-.13$ ).

Assessing the fitness of the 2-factor structure by Tsui & Kennedy (2009) was accomplished with marginal fit. The goodness-of-fit indexes included:  $\chi(53) = 109.45$ ,  $p < .001$ ;  $CMIN/DF = 2.07$ ,  $CFI = .90$ ,  $NFI = .82$ ,  $TLI = .85$  and  $RMSEA = .10$  (90% CI of  $RMSEA = .07-.12$ ). The factor loadings for *efficacy in classroom management* (Items 1, 2, 6, 7, 8) ranged from .54 to .77 and for *efficacy for teaching and support* (Items 3, 4, 5, 9, 10, 11, 12), from .58 to .77. The two models (Cheung, 2006; Tsui & Kennedy, 2009) were almost identical. The only difference between these two models was that Item 2 was moved from the *teaching and support* factor to the *classroom management* factor.

As the assumption of normality was not violated, and no outliers were found, the means and standard deviations of the three factors were obtained. A series of one sample t-tests or ANOVAs found no significant differences between the two factors for the three independent variables (Table 4).

#### Revising the 2-factor structure to a 9-item form

Since the CFA suggested a poor fit of the factor models proposed by Cheung (2006) and Tsui & Kennedy (2009), EFA was conducted to examine latent structure of the 12 TSES items using the current sample. Following the procedures described by Pallant (2016), oblique rotations were performed to reveal the simplest structure. Oblique rotation, assuming the correlated rotated factor, was preferred because it was more realistic and accurate to obtain meaningful underlying factors (Tabachnick & Fidell, 2014).

Two factors were disclosed with eigenvalues larger than 1, which accounted for 56.64% of the variance. The Parallel Analysis (undertaken using MonteCarlo PCA for Parallel Analysis, Watkins, 2000), however, indicated only one factor with eigenvalues exceeding the corresponding criterion values for a randomly generated data matrix of the same size (12 variables  $\times$  120 participants). To interpret the result of the Parallel Analysis, the actual eigenvalues generated by SPSS were compared to the corresponding criterion values obtained from parallel analysis (Hayton *et al.*, 2004, p. 194). The first actual eigenvalue is compared to the first parallel average random eigenvalues, the second actual eigenvalue to second parallel average random eigenvalue, etc. When the actual eigen-

**Table 3. Model comparison, goodness of fit indices.**

Model	CM/df	NFI	TLI	CFI	RMSEA
<b>Good Model Fit</b>	<3	> .95	> .95	> .95	< .07
<b>Acceptable Model Fit</b>	<5	> .90	> .90	> .90	.07-.10
1. One-factor model (12 items)	2.34	.80	.81	.87	.11
2. One-factor model (11 items)	3.64	.71	.66	.76	.12
3. One-factor model (10 items)	3.38	.75	.77	.81	.14
4. Two-factor model (9 items)	1.57	.91	.94	.96	.07
5. Two-factor model (Tsui & Kennedy, 2009) (12 items)	2.07	.82	.85	.90	.10
6. Two-factor model (Cheung, 2006) (12 items)	2.24	.81	.82	.88	.10
7. Two-factor model (Cheung, 2006) (11 items)	2.23	.90	.83	.84	.10
8. Three-factor model (Tschannen & Hoy, 2001) (12 items)	2.09	.83	.84	.90	.10

**Table 4. Descriptive statistics based on the 12-item two factor solution.**

Gender		N	Mean	SD
Classroom Management	Male	36	6.61	1.05
	Female	83	6.62	0.91
Teaching and Learning	Male	36	6.60	0.77
	Female	83	6.40	0.86
Overall Efficacy	Male	36	6.61	0.80
	Female	83	6.47	0.83
Age Group		N	Mean	SD
Classroom Management	≤ 25 years old	50	6.53	0.96
	≥ 26 years old	70	6.67	0.94
Teaching and Learning	≤ 25 years old	50	6.40	0.77
	≥ 26 years old	70	6.49	0.89
Overall Efficacy	≤ 25 years old	50	6.44	0.79
	≥ 26 years old	70	6.55	0.85
Teaching Level		N	Mean	SD
Classroom Management	Kindergarten	43	6.78	0.84
	Primary	58	6.58	1.03
	Secondary	18	6.33	0.88
	Total	119	6.62	0.95
Teaching and Learning	Kindergarten	43	6.48	0.78
	Primary	58	6.47	0.89
	Secondary	18	6.41	0.77
	Total	119	6.47	0.83
Overall Efficacy	Kindergarten	43	6.58	0.76
	Primary	58	6.51	0.89
	Secondary	18	6.38	0.71
	Total	119	6.52	0.82

value was larger than the parallel average random eigenvalue, the factor was retained. In the present study, the first actual eigenvalue from the PCA (i.e. 5.429) was larger than its parallel criterion value (i.e. 1.563), therefore the first component was retained. The second actual eigenvalue (i.e. 1.368), however, was smaller than its parallel criterion value (i.e. 1.383), and therefore the second component was rejected. Hence, the parallel study suggested a one-factor solution.

The one-factor solution from Maximum Likelihood extraction explained a total of 45.24% of variance. The CFA indicated a poor fit:  $\chi(54)=126.24$ ,  $p < .001$ ; CMIN/DF= 2.34, CFI= .87, NFI= .80, TLI= .81 and RMSEA= .11 (90% CI of RMSEA= .08-.13). In order to refine the model to achieve the acceptable model fit, item deletion was considered. Under EFA, an item with communality below 0.4 should be removed (Henson *et al.*, 2004). Under CFA, an item with factor loading below .5 was considered as a bad item (Hair *et al.*, 2010). Under EFA, the communalities of items 2 ( $h^2 = .27$ ), 5 ( $h^2 = .29$ ) and

9 ( $h^2 = .35$ ) were below 0.4. Under CFA, the factor loading of Item 2 was .49. Therefore, Item 2 was deleted and EFA as well as CFA were rerun. Table 5 indicated that Items 9 and 5 were deleted one by one subsequently in the next two rounds of analyses (i.e. Model 3 & Model 4).

In the 4<sup>th</sup> round of analysis, after deleting Items 2, 9 and 5, the same procedure was carried out on the remaining nine items. The two-factor solution explained 63.0% of the total variance, with Factors 1 to 2 contributing 51.60% and 11.40% of the variance respectively. The commonality ranged from .55 to .75. The first factor *classroom management* composed of Items 1, 3, 6, 7, 8 and the second factor *teaching and learning*, Items 4, 10, 11, 12. The CFA revealed a good fit:  $\chi(46)=40.86$ ,  $p = .03$ ; CMIN/DF= 1.57, CFI= .96, NFI= .91, TLI= .94 and RMSEA= .07 (90% CI of RMSEA= .02-.11). The loadings for the first factor ranged from .66 to .77; the loadings for the second factor ranged from .64 to .77 (Table 6).

To identify the best models among the models from the present study, Cheung (2006), Tsui & Kennedy (2009), and Tschannen & Hoy (2001), model comparison with AIC was used (Burnham & Anderson, 2004, p. 270–271). It was found that the 9-item model with the lowest AIC value (Model 4) outperformed the others (Table 7).

Table 8 represents the figures of Composite Reliability (CR), Average Variance Extracted (AVE), Maximum Shared Variance (MSV) and Average Shared Variance (ASV). The first two are the indicators of convergent validity and the last two, discriminant validity (Hair *et al.*, 2010; Gaskin, 2012). To achieve convergent validity, three conditions must be met: (1) CR for each factor should be at least .70; (2) AVE should be at least .50; and (3) CR should be larger than AVE. To establish discriminant validity, AVE should be larger than corresponding MSV and ASV. Therefore, both factors have satisfactory convergent validity (i.e., CR values for both factors are larger than .70; AVE values of both factors are at least .5; and CR values are larger than their corresponding AVE values) even though they were not well distinguished (i.e. MSV and ASV for both factors are larger than their corresponding AVE values). Furthermore, the overall and factor reliability of this 9-item form were satisfactory, with .88 (overall), .84 (*classroom management*) and .78 (*teaching and learning*) respectively).

Mean comparisons were conducted with the 9-item revised TSE-C two factor solution. Similar to the findings when using the two-factor structure of Cheung (2006), no significant differences of means were found for the three independent variables. The response scale and the descriptive statistics of the 9-item factor solution are presented in Table 9.

## Discussion

Teacher self-efficacy has been found to be a strong determinant in enabling effective inclusive education (Holzberger *et al.*, 2013; Küsting *et al.*, 2016). Yet in Macao, teachers' perceptions of self-efficacy regarding their involvement with inclusive education have to date not been very positive (Cheung *et al.*, 2015; HKIEd, 2012). Teacher preparation programs have recently been



**Table 5. Model comparison, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).**

	Model							
	1	2	3	4	5	6	7	8
Range of commonality (EFA)	.27-.60	.31-.60	.34-.60	.55-.75	--	--	--	--
Item with the lowest commonality	2	9	5	10	--	--	--	--
Range of factor loadings (CFA)	.49-.75	.45-.72	.52-.72	.64-.77	.54-.77	.46-.77	.53-.74	.48-.74
Item with the lowest loading	2	9	5	10	2	2	9	2
AIC	198.28	229.23	180.93	96.86	183.45	192.88	163.91	184.47

**Table 6. Exploratory factor analysis of the 9-item factor solution of the present study.**

	M	SD	$h^2$	Item	CM	TL	CFA loading
1	6.26	1.23	.68	How much can you do to control disruptive behavior in the classroom?	.88		.66
3	6.23	1.38	.63	How much can you do to get students to believe they can do well in schoolwork?	.74		.71
6	6.66	1.25	.75	How much can you do to get children to follow classroom rules?	.86		.77
8	6.48	1.25	.59	How well can you establish a classroom management system with each group of students?	.48		.72
4	6.71	1.10	.67	How much can you do to help your students value learning?		.80	.77
7	7.01	1.05	.58	How much can you do to calm a student who is disruptive or noisy?		.53	.71
10	6.77	1.11	.55	To what extent can you provide an alternative explanation or example when students are confused?		.77	.64
11	6.03	1.63	.61	How much can you assist families in helping their children do well in school?		.81	.68
12	6.41	1.12	.62	How well can you implement alternative strategies in your classroom?		.76	.73

Note. CM = efficacy for/in classroom management; TL = efficacy in teaching and learning; CFA = confirmatory factor analysis.

changed in Macao to better meet the needs of teachers working in inclusive schools (Monteiro & Forlin, in press), although there is no current validated measure to monitor the outcome of this over time. Ensuring a validated scale for measuring teacher

self-efficacy is essential as major changes in practice will continue the movement tending towards greater inclusion of children with disabilities in regular schools in Macao. The aim of this study, therefore, was to review the use of the long and short

**Table 7. Comparison of underlying dimensions of teacher sense of efficacy derived from confirmatory factor analysis in different studies.**

Item	Mean	SD	Present Study	Tschannen & Hoy (2001)	Cheung, (2006)	Tsui & Kennedy (2009)
5	6.59	1.15	Deleted	IS (.61)	TL (.58)	TS (.59)
9	6.42	1.19	Deleted	IS (.55)	TL (.52)	TS (.53)
10	6.78	1.10	TL (.64)	IS (.64)	TL (.63)	TS (.63)
12	6.43	1.12	TL (.73)	IS (.74)	TL (.72)	TS (.73)
3	6.23	1.38	CM (.71)	SE (.62)	TL (.59)	TS (.58)
4	6.72	1.10	TL (.77)	SE (.70)	TL (.77)	TS (.77)
11	6.03	1.62	TL (.68)	SE (.63)	TL (.68)	TS (.59)
2	6.44	1.31	Deleted	SE (.48)	TL (.46)	CM (.54)
1	6.27	1.23	CM (.66)	CM (.63)	CM (.64)	CM (.66)
6	6.67	1.25	CM (.77)	CM (.74)	CM (.73)	CM (.73)
7	7.02	1.05	CM (.71)	CM (.72)	CM (.72)	CM (.72)
8	6.48	1.24	CM (.72)	CM (.74)	CM (.74)	CM (.74)
Model Fit			Good	Acceptable	Poor	Marginal
Cronbach's Alpha			TL: .78 CM: .84 Total: .88	IS: .67 SE: .69 CM: .80 Total: .88	TL: .82 CM: .80 Total: .88	TS: .83 CM: .81 Total: .88

Note: ( ) = the CFA factor loadings; SE = efficacy for student engagement; CM = efficacy for/in classroom management; IS = efficacy for instructional strategies; TS = efficacy for teaching and support; TL = efficacy in teaching and learning.

**Table 8. Reliability and validity of Model 4.**

Model 4	Items	CR	AVE	MSV	ASV	Cronbach's Alpha
Classroom Management (CM)	1, 3, 6, 7, 8	0.84	0.51	0.69	0.69	0.84
Teaching and Learning (T&L)	4, 10, 11, 12	0.80	0.50	0.69	0.69	0.78 0.88 (Total)

versions of the TSES and the different factor structures in a Chinese format to assess the suitability for monitoring long term changes in teacher self-efficacy for inclusive education.

Administering the long version in Chinese demonstrated very good reliabilities (Cronbach's alpha >.810), therefore, validating its use in a Chinese format. Applying confirmatory factor analysis on the 12-item Chinese version developed by Kennedy & Hui (2004) and further validated by Cheung (2006), the two-factor structure was, however, not fully supported. Exploratory and confirmatory analysis identified a stronger 9-item, two factor scale. In this new scale, the *classroom management* factor (n=5) was identical to previous studies. The *teaching and learning* factor (n=4), though, had the most discrepancy. With the deletion of Items 2, 5, and 9, the remaining four items were

grouped under the second factor in both the new 9-item and the prior 12-item factor structures. Psychometrically, nevertheless, the level of internal consistency (Cronbach's Alpha = 0.88) of these two versions was approximately the same. This implied that the length of the scale is as important as item intercorrelation.

The two factor solution obtained in the present study is similar to previous factor analysis of TSES involving Hong Kong samples (e.g. Kennedy & Hui, 2004; Cheung, 2006) and items grouped under the two factors are different. For example, a recent factor analytic study (Wu & Chim, 2017) of the 12-item TSES on the sample of 609 Hong Kong in-service primary and secondary school teachers also revealed a two factor solution identical to the one obtained by Cheung (2006) but slightly different from the one discovered in the present study. The class management factor from the present study consisted of five items (Items 1, 6, 7, 8, 3), whereas, the one by Cheung (2006) and Wu & Chim (2017) included four items (1, 6, 7, 8). Item 3 "How much can you do to get students to believe they can do well in schoolwork?" was perceived by Macau teachers as a class management (or discipline) rather than teaching and learning issue. The two factors in the present study were, nonetheless, not well differentiated. This is consistent to Cheung (2008)'s comparative study of teacher efficacy on Hong Kong and Shanghai primary in-service teachers. When EFA was conducted separately on the Hong Kong sample (n=725), one-factor solution was extracted.

**Table 9. Descriptive statistics based on the 9-item two factor solution (present study).**

Gender		N	Mean	SD
Classroom Management	Male	36	6.59	1.03
	Female	83	6.52	0.94
Teaching and Learning	Male	36	6.65	0.82
	Female	83	6.43	1.03
Overall Efficacy	Male	36	6.67	0.86
	Female	83	6.48	0.90
Age Group		N	Mean	SD
Classroom Management	≤ 25 years old	50	6.44	0.97
	≥ 26 years old	70	6.60	0.95
Teaching and Learning	≤ 25 years old	50	6.44	0.93
	≥ 26 years old	70	6.52	1.01
Overall Efficacy	≤ 25 years old	50	6.44	0.87
	≥ 26 years old	70	6.56	0.89
Teaching Level		N	Mean	SD
Classroom Management	Kindergarten	43	6.73	0.89
	Primary	58	6.49	1.08
	Secondary	18	6.27	0.79
	Total	119	6.54	0.97
Teaching and Learning	Kindergarten	43	6.50	0.84
	Primary	58	6.53	1.06
	Secondary	18	6.40	0.86
	Total	119	6.50	0.97
Overall Efficacy	Kindergarten	43	6.63	0.82
	Primary	58	6.50	0.97
	Secondary	18	6.33	0.69
	Total	119	6.52	0.88

It is central to acknowledge that the data set used for this validation purpose was relatively small ( $n=118$ ), compared to that of Tsui & Kennedy (2009) ( $n=173$ ) but larger than that of Cheung (2006) ( $n=71$ ). To confirm the validity of the new version of the scale considerably more data needs to be collected from a much larger sample, as these differences may in part be due to the different sample sizes and potentially different interpretations of the items. According to Lynch (2013), the sample size must be proportionate to the total overall population. While obtaining an accurate measure of suitable size can be quite difficult, it would seem clear that the relatively small number of participants used for identifying efficacy in this study will not yield enough rigor to enable any generalization.

Although it is very tempting to modify an original scale and use a shorter version, as can be seen in these results, this can be problematic due to the inconsistent outcomes. By analyzing the original 24-item and 12-item scales with three factors, and then

the 12-item Chinese scale with the two factors, and finally the new 9-item scale with two factors; different results were found. Although no significant differences were evident for age, or teaching level, across all versions, there was a significant difference noted for gender when using the long 24-item total scale. In this instance a moderate effect size was found with men indicating higher efficacy in applying instructional strategies than women. Application of the long version of the scale with 103 primary school teachers in China, similarly found gender differences with instructional strategies, although in a converse way with females indicating higher levels of efficacy than male teachers (Manzar-Abbas & Liu, 2015). Another very large study in China of 1027 special education teachers also found female teachers rated themselves significantly higher on efficacy on the long 24-item scale (Minghui *et al.*, 2018). Additional research by Cheung (2008), used the HK-TSE 12-item scale (Kennedy & Hui, 2006) with Chinese and Hong Kong primary teachers. The numbers in these studies were large ( $> 575$ ), but Cheung still had to use two different versions of the scale due to inconsistencies with the item structures finalized upon different cultural differences. In addition, psychometric analysis of the scales in the two different regions found different factors with different loadings. This by itself is not of major concern, but it does indicate that greater evidence is needed across a wider sample of Chinese speakers from different cultural backgrounds, to ensure the validity, reliability and particularly the generalizability of the scales.

It must also be remembered that using the TSES requires teachers to self-rate their own efficacy as a teacher. Such self-identification is not always found to be reliable as there are many reasons why teachers may want to present themselves as more capable than they are. As found by Poulou *et al.*, (2018), in their recent use of the TSES in Greece, significant differences were found by neutral observers' ratings compared to those given by the teachers themselves. In the Asian culture this is particularly pertinent as retaining a teaching position is decidedly competitive and teachers need to demonstrate their teaching skills as highly effective. The potential tendency to overrate teaching self-efficacy needs to be explored further. Additional consideration also needs to be given to who is administering the questionnaire. According to Lüke & Grosche (2018), a key variable in explaining differences in teacher perceptions has been associated with teachers' sensitivities to the attitudes of the university conducting the research. Lüke and Grosche identify this as social desirability-induced validity problems and posit that the likely impact of this on research on attitudes towards inclusion has been ignored.

## Conclusion

For the purpose of identifying the perceptions of pre-service teachers in Macao regarding their self-efficacy beliefs, the conclusions we draw are limited. From the evidence presented here it is apparent that while applying a quick and easy scale to identify teacher self-efficacy may be considered useful, there are far too many cultural, demographic, and edifying issues that may prevent any generalizable conclusions being drawn. The

use of a measure such as the TSES, TSES-C, or the revised 9-item scale, must be applied with caution and should be used alongside qualitative data that can provide greater depth into the potential reasons underpinning teachers' self-perceptions of teaching efficacy. While the revised 9-item scale gave high reliabilities for this small group of participants, considerable further research is needed with larger numbers to confirm generalizability of this.

Accountability and quality assurance are key dimensions in the 21<sup>st</sup> century for authenticating changes in education. Governments, training institutions, and schools need to be responsible for ensuring that teachers are provided with suitable training and relevant ongoing support so that they feel confident in being able to provide appropriate educational experiences for the increasing diversity of students being included in regular schools. If teachers in Macao (and elsewhere) are to embrace Government policy changes towards greater inclusive education, then identifying and managing their self-efficacy in being able to implement this change will be critical to its success. A valid on-going way of measuring this will be important to ensure appropriate support relevant to improving teacher self-efficacy for inclusion is provided, to prepare them for this significant change.

## Data availability

### Underlying data

Open Science Framework: Teachers' Sense of Efficacy Scale (TSES) Macau Study Dataset, <https://doi.org/10.17605/OSF.IO/3PX6J> (Monteiro, 2020a).

### Extended data

Open Science Framework: Teachers' Sense of Efficacy Scale (TSES) Short Version Chinese, <https://doi.org/10.17605/OSF.IO/T89HQ> (Monteiro, 2020b).

This project contains the following extended data:

- New 12-item version of TSES in Chinese
- New 9-item version of TSES in Chinese

Data are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

## Acknowledgments

The authors are grateful to Edmund Chan for his support with the data analysis.

## References

- Arbuckle JL: **Amos**. (Version 24.0) [Computer Program]. Chicago: SPSS. 2018.
- Armor D, Conroy-Oseguera P, Cox M, *et al.*: **Analysis of the school preferred reading programs in selected Los Angeles minority schools**. (Report No. R-2007-LAUDS). Santa Monica, CA : Rand Corporation (ERIC Document Reproduction Service No. 130 243). 1976.  
[Reference Source](#)
- Ashton PT, Olejnik S, Crocker L, *et al.*: **Measurement problems in the study of teachers' sense of efficacy**. Paper presented at the annual meeting of the American Educational Research Association, New York. 1982.  
[Reference Source](#)
- Bandura A: **Self-efficacy: The exercise of control**. New York: W. H. Freeman and Company. 1977.  
[Reference Source](#)
- Bandura A: **Self-efficacy: Toward a unifying theory of behavioral change**. *Psychol Rev.* 1977; **84**: 191–215.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Berman P, McLaughlin M, Bass G, *et al.*: **Federal programs supporting educational change. Factors effecting implementation and continuation, Vol. 7**. Santa Monica, CA : The Rand Corporation. 1977.  
[Reference Source](#)
- Burnham KP, Anderson DR: **Multimodel inference: Understanding AIC and BIC in model selection**. *Sociological Methods & Research.* 2004; **33**(2): 261–304.  
[Publisher Full Text](#)
- Byrne BM: **Structural equation modeling with AMOs: Basic, concepts, applications, and programming**. (2<sup>nd</sup> ed.). New York: Routledge. 2010.  
[Reference Source](#)
- Chao G, Forlin C, Ho FC: **Improving teaching self-efficacy for teachers in inclusive classrooms in Hong Kong**. *International Journal of Inclusive Education.* 2016; **20**(11): 1142–1154.  
[Publisher Full Text](#)
- Cheung HY: **The measurement of teacher efficacy: Hong Kong primary in-service teachers**. *Journal of Education for Teaching.* 2006; **32**(4): 435–451.  
[Publisher Full Text](#)
- Cheung HY: **Validation of the Chinese version of Teachers' Sense of Efficacy Scale for Macao pre-service teachers**. *Pacific-Asian Education Journal.* 2006b; **18**(1): 22–31.  
[Reference Source](#)
- Cheung HY: **Teacher efficacy: A comparative study of Hong Kong and Shanghai primary in-service teachers**. *Aust Educ Res.* 2008; **35**(1): 103–123.  
[Publisher Full Text](#)
- Cheung H, Wu J, Hui S: **Chinese attitudes toward inclusive education: Perspectives of Hong Kong and Macao secondary school teachers, students and parents**. *International Research Journal for Quality in Education.* 2015; **2**(1): 1–14.  
[Reference Source](#)
- Chong S, Forlin C, Au ML: **The influence of an inclusive education course on attitude change of pre-service secondary teachers in Hong Kong**. *Asia Pacific Journal of Teacher Education.* 2007; **35**(2): 161–179.  
[Publisher Full Text](#)
- Copfer S, Specht J: **Measuring effective teacher preparation for inclusion**. In C. Forlin, & T. Loreman (Eds.), *Measuring inclusive education*. United Kingdom: Emerald Group Publishing Limited. 2014; **3**: 93–113.  
[Reference Source](#)
- Correia A, Monteiro E, Teixeira V, *et al.*: **The interplay between a Confucian-heritage culture and teachers' sentiments and attitudes towards inclusion in Macao**. *Eur J Spec Needs Educ.* 2019; **5**(2).  
[Reference Source](#)
- Direcção dos Serviços de Educação e Juventude (DSEJ): **Alteração ao "Regime educativo especial" - documento de consulta**. Macao: Região Administrativa Especial de Macao [Alteration to the Special Education Regime-consultation document]. 2015.  
[Reference Source](#)
- Direcção dos Serviços de Educação e Juventude (DSEJ): **Curriculum framework for formal education of local education system**. 2014.  
[Reference Source](#)
- Dragan D, Topšek D: **Introduction to structural equation modeling: Review, methodology and practical applications**. Paper presented at the International Conference on Logistics & Sustainable Transport 2014, Celje, Slovenia. 2014.  
[Reference Source](#)
- Duffin LC, French BF, Patrick H: **The Teachers' Sense of Efficacy Scale: Confirming the factor structure with beginning pre-service teachers**. *Teach Teach Edu.* 2012; **28**(6): 827–834.  
[Publisher Full Text](#)
- Forlin C, Loreman T, Sharma U: **A system-wide professional learning approach about inclusion for teachers in Hong Kong**. *Asia-Pac J Teach Edu.* 2014; **42**(3): 247–260.  
[Publisher Full Text](#)

- Forlin C, Chambers D: **Teacher preparation for inclusive education: Increasing knowledge but raising concerns.** *Asia-Pac J Teach Edu.* 2010; **39**(1): 17–32. [Publisher Full Text](#)
- Gaskin J: **Validity and reliability.** Gaskination's StatWiki. 2012. [Reference Source](#)
- Gibson S, Dembo MH: **Teacher efficacy: A construct validation.** *Journal of Educational Psychology.* 1984; **78**(4): 569–582. [Reference Source](#)
- Guo Y, Piasta SB, Justice LM, et al.: **Relationships among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains.** *Teaching and Teacher Education.* 2010; **26**(4): 1094–1103. [Publisher Full Text](#)
- Hair JF, Black WC, Babin BJ, et al.: **Multivariate data analysis: A global perspective.** (7<sup>th</sup> ed.). New York: Pearson. 2010. [Reference Source](#)
- Hayton JC, Allen DG, Scarpello V: **Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis.** *Organ Res Methods.* 2004; **7**(2): 191–205. [Publisher Full Text](#)
- Henson RK, Capraro RM, Capraro MM: **Reporting practices and use of exploratory factor analyses in educational research journals: Errors and explanation.** *Res Sch.* 2004; **11**(2): 61–72. [Reference Source](#)
- HKIEd: **Macao special education report.** Hong Kong: The Hong Kong Institute of Education - Special Learning needs and Inclusive Education Centre. 2012.
- Holzberger D, Philipp A, Kunter M: **How teachers' self-efficacy is related to instructional quality: A longitudinal analysis.** *J Educ Psychol.* 2013; **105**(3): 774. [Publisher Full Text](#)
- Hui SKF, Kennedy KJ, Cheung HY: **Hong Kong and Macao pre-service teachers' sense of efficacy: A cross cultural investigation using the Chinese version of the 12-item TSE (C-TSE).** *The Asia-Pacific Education Research.* 2006; **15**: 41–62. [Reference Source](#)
- Kennedy KJ, Hui SKF: **Self-efficacy as a key attribute for curriculum leaders: Linking individual and organizational capacity to meet the challenges of curriculum reform.** Paper presented at the Annual Conference of the Commonwealth Council of Educational Administration and Management, Hong Kong. 2004.
- Kennedy KJ, Hui SKF: **Developing teacher leaders to facilitate Hong Kong curriculum reforms: Self-efficacy as a measure of teacher growth.** *Int J Educ Reform.* 2006; **15**(1): 114–128. [Publisher Full Text](#)
- Klassen RM, Bong M, Usher EL, et al.: **Exploring the validity of a teachers' self-efficacy scale in five countries.** *Contemp Educ Psychol.* 2009; **34**(1): 67–76. [Publisher Full Text](#)
- Künsting J, Neuber V, Lipowsky F: **Teacher self-efficacy as a long-term predictor of instructional quality in the classroom.** *Eur J Psychol Educ.* 2016; **31**(3): 299–322. [Publisher Full Text](#)
- Loerman T, Sharma U, Forlin C: **Do pre-service teachers feel ready to teach in inclusive classrooms?** *Aust J Teach Edu.* 2013; **38**(1): 24–44. [Publisher Full Text](#)
- Lüke T, Grosche M: **What do I think about inclusive education? It depends on who is asking. Experimental evidence for a social desirability bias in attitudes towards inclusion.** *Int J Incl Educ.* 2018; **22**(1): 38–53. [Publisher Full Text](#)
- Lynch SM: **Using statistics in social research.** Springer: New York. 2013.
- Main S, Hammond L: **Best practice or most practiced? Pre-service teachers' beliefs about effective behaviour management strategies and reported self-efficacy.** *Aust J Teach Edu.* 2008; **33**(4): 28–39. [Publisher Full Text](#)
- Manzar-Abbas SS, Liu L: **Self-efficacy beliefs of Chinese primary school Teachers.** *Pak J Psychol Res.* 2015; **30**(2): 289–303. [Reference Source](#)
- Minghui L, Lei H, Xiaomeng C, et al.: **Teacher efficacy, work engagement, and social support among Chinese special education school teachers.** *Front Psychol.* 2018; **9**: 648. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Monteiro E: **Teachers' Sense of Efficacy Scale (TSES) Macau Study Dataset.** 2020a. <http://www.doi.org/10.17605/OSF.IO/3PX6J>
- Monteiro E: **Teachers' Sense of Efficacy Scale (TSES) Short Version Chinese.** 2020b. <http://www.doi.org/10.17605/OSF.IO/T89HQ>
- Monteiro E, Forlin C: **Enhancing teacher education by utilizing a revised PGDE curriculum as a fundamental resource for inclusive practices in Macao.** In: Loreman, T., Goldan, J., & Lambrecht, J., (Eds.); *Resourcing inclusive education.* International perspectives on inclusive education, UK: Emerald. (in press); **15**.
- Monteiro E, Correia A, Forlin C, et al.: **Perceived efficacy of teachers in Macao and their alacrity to engage with inclusive education.** *Int J Incl Educ.* 2018 ; 1464-5173 (Online). [Publisher Full Text](#)
- Osborne JW: **Best practices in exploratory factor analysis.** Scotts Valley, CA: CreateSpace Independent Publishing. 2014. [Reference Source](#)
- Osborne JW, Costello AB, Kellow JT: **Best practices in exploratory factor analysis.** In: *Best practices in quantitative methods.* Thousand Oaks, CA SAGE. 2008; 87–99. [Publisher Full Text](#)
- Pallant J: **SPSS survival manual: A step by step guide to data analysis using IBM SPSS.** 5th ed. Maidenhead, Berkshire, England: McGraw Hill. 2016. [Reference Source](#)
- Poulou MS, Redd LA, Dudek YCM: **Relation of teacher self-efficacy and classroom practices: A preliminary investigation.** *Sch Psychol Int.* 2018; **40**(1): 25–48. [Publisher Full Text](#)
- Rose JS, Medway FJ: **Measurement of teachers' beliefs in their control over student outcome.** *J Educ Res.* 1981; **74**(3): 185–190. [Publisher Full Text](#)
- Rotter JB: **Generalised expectancies for internal versus external control of reinforcement.** *Psychol Monogr.* 1966; **80**(1): 1–28. [PubMed Abstract](#) | [Publisher Full Text](#)
- Singh P, Rowan L, Allen J: **Reflection, research and teacher education.** *Asia-Pacific Journal of Teacher Education.* 2019; **47**(5): 455–459. [Publisher Full Text](#)
- Sivan A, Chan DWK, Kwan YW: **Psychometric evaluation of the Chinese version of the questionnaire on teacher interaction (C-QTI) in Hong Kong.** *Psychol Rep.* 2014; **114**(3): 823–842. [PubMed Abstract](#) | [Publisher Full Text](#)
- Tabachnick BG, Fidell LS: **Using multivariate statistics.** 6th ed. Boston: Pearson Education. 2014. [Reference Source](#)
- Teixeira V, Correia A, Forlin C, et al.: **Placement, inclusion, law and teachers' perceptions in Macao's schools.** *Int J Incl Educ.* 2018; **22**(9): 1014–1032. [Publisher Full Text](#)
- Tschannen-Moran M, Woolfolk Hoy A: **Teacher efficacy: Capturing an elusive construct.** *Teaching and Teacher Education.* 2001; **17**(7): 783–805. [Publisher Full Text](#)
- Tschannen-Moran M, Woolfolk Hoy A: **The differential antecedents of self-efficacy beliefs of novice and experienced teachers.** *Teaching and Teacher Education.* 2006; **23**(6): 944–956. [Publisher Full Text](#)
- Tschannen-Moran M, Hoy AW, Hoy WK: **Teacher efficacy: Its meaning and measure.** *Rev Educ Res.* 1998; **68**(2): 202–248. [Publisher Full Text](#)
- Tsigilis N, Koustelios A, Grammatikopoulos V: **Psychometric properties of the Teachers' Sense of Efficacy Scale within the Greek educational context.** *J Psychoeduc Assess.* 2010; **28**(2): 153–162. [Publisher Full Text](#)
- Tsui KT, Kennedy KJ: **Evaluating the Chinese version of the Teacher Sense of a Efficacy Scale (C-TSE): Translation adequacy and factor structure** *The Asia-Pacific Education Researcher.* 2009; **18**(2): 245–260. [Publisher Full Text](#)
- Watkins MW: **MonteCarlo PCA for parallel analysis [computer software].** State College. PA Ed & Psych Associates. 2000. [Reference Source](#)
- Wolf E, Harrington K, Clark S, et al.: **Sample size requirements for structural equation models: An Evaluation of Power, Bias, and Solution Propriety.** *Educ Psychol Meas.* 2013; **73**(6): 913–934. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wu L, Chim HY: **The short-form Teacher Efficacy Scale: A study of reliability and validity.** *Psychology: Techniques and Applications.* 2017; **5**(11): 672–679.



# Open Peer Review

Current Peer Review Status:  

---

## Version 1

Reviewer Report 08 December 2020

<https://doi.org/10.21956/emeraldopenres.14612.r27221>

© 2020 Martin S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Stewart Martin** 

<sup>1</sup> University of Hull, Kingston upon Hull, United Kingdom

<sup>2</sup> University of Saint Joseph, Macau SAR, Peoples Republic Of China

The article seeks to validate two versions of the Teachers' Sense of Efficacy Scale with pre-service teachers in Macau and also seeks to explore the impact of culture, demographics and other issues that may impact on the usefulness of such scales.

As with all educational research instruments, there is a need to explore and establish the value of those proposing to be able to accurately measure constructs such as self-efficacy and practices such as inclusivity and to examine the implications of their application in a wide range of different contexts. The authors are commended for their interest and research in these important and relevant areas.

### Evidence and assumptions

The paper makes use of some statements for which evidence is lacking and in some cases these make implicit assumptions, such as (p.3) "*teacher self-efficacy has proven to play an essential role in improving teachers' pedagogical practices*"; and "*Teacher self-efficacy has shown to be a salient variable in predicting teachers' capability to execute effective practices and transformation within educational reforms*". These statements make claims about causal relationships between teachers' beliefs and improvements in teachers' practices or educational transformations as though these have been established as matters of fact when this is not the case.

From time to time we have all met individuals who believe that they are capable of doing certain things, but this does not always mean that their beliefs represent what is true. I may hold many sincerely-felt convictions that I am highly competent to do a number of things and people who know me may even agree with me – but none of this is any kind of guarantee that my perceptions are correct and are reflected by reality. We would normally require evidence of actual competence before accepting that an individual claiming to be a competent teacher in any particular field of their practice was in fact as competent as they believe. The relationship between self-belief and performance is complex and, even if we were able to demonstrate a direct causal relationship

between these, it would be extremely difficult to know which of these was the cause and which was the effect or the degree to which this was the case. There is a great deal more to being a highly effective teacher, surgeon, pilot, engineer or builder than simply self-belief. In all such cases we would be cautious of accepting that the level of an individual's beliefs were a meaningful and strong indicator that those directly create a certain level of professional conduct, practice and skill.

The article does not sufficiently reflect such caution. One example is seen in the claim that "By developing an effective and valid measure of self efficacy it will be possible to measure teachers' perceived readiness for inclusive education over time; *to safeguard that they are ready to become inclusive practitioners.*"(my emphasis). Although this statement from the introduction helpfully clarifies some of the impetus behind producing the article, it also identifies what appears to be a tacit assumption throughout the paper; that teachers' statements about their self-efficacy (even if their beliefs are well-founded) *cause* them to become more inclusive practitioners who teach an effective curriculum. Such claims are highly unlikely to be true and to help remove any unintentional lack of clarity about this, the paper would benefit from a deeper and more nuanced exploration of the nature and identification of causality in educational research and the role of self-report data when seeking to explain in valid and reliable ways the relationship of particular variables to specific educational outcomes.

Points that must be addressed to make the article more scientifically sound:

1. Amend the article text to remove implicit or explicit claims or suggestions of causality that can be deduced from your data or that are purported to result from levels of self-efficacy or any measures of this.
2. Where relevant emphasise more strongly throughout the discussion what can and cannot be deduced from measures of association.
3. Include a deeper and more nuanced exploration of the nature and means of identifying causality in educational research and the role of self-report data when seeking to explain in valid and reliable ways the relationship of particular variables to specific educational outcomes.

**An effective curriculum**

What comprises an effective curriculum in the context of inclusive education is a very interesting question but one that is difficult to address or answer due to its inherent conceptual vagueness. Is the 'curriculum' simply the factual content of the taught classes in an institution or should this also include any institutional or socially embedded values, behaviours, ethics, beliefs and morals? Does the provision of an inclusive education depend upon the actions of other social agents or individuals beyond those with formal educational purposes? Inclusivity in education would seem to presume the active involvement of all of these to some degree but how should these be represented and under such a scenario what would we accept as 'effective' and how would this be agreed, understood, validated, recognised or made known?

Although this may have been outside the initial scope of the present study, I strongly encourage the authors to address the matter of how teachers actually *perform* 'inclusive education' and exactly what this term means. It would also be most interesting to see the authors explore the role played by 'efficacy' in the curriculum and its associated assessment strategies – and what an 'inclusive curriculum' would look like and could be shown to require in order to be classified as *effective*. Such enquiry may help to shed more light on whether inclusive education is meaningfully

mediated by any changes that have or could be made to teacher education or by any beliefs teachers may have or may acquire regarding their professional efficacy.

Points that must be addressed to make the article more scientifically sound:

1. Clarify what is meant by 'inclusive education'. Discuss also whether there is a clear and widely accepted definition of inclusive education. This is important to address because any instruments seeking to measure an individual's ability to provide inclusive education must be founded on a standardised and widely agreed definition of what 'inclusion' is and how it may be recognised.
2. Are there uniform and widely agreed *objectives* for inclusive education and if so, what specific outputs will be produced and how can these outputs be reliably measured?
3. Clarify what is meant by an 'inclusive curriculum'. Explain the ways in which its features would differ from, or be similar to, any other kind of curriculum.
4. Does robust, valid and reliable measures for 'inclusion' and 'inclusivity' exist? If so, explain what these are and how they work.
5. What specific knowledge, skills or competencies have been reliably shown to *directly facilitate* inclusive education?
6. Do any of the instruments or concepts discussed in the article represent *valid and reliable* measures of any such individual knowledge, skills or competencies that may be possessed by an individual? If so, where and what is the evidence for this and, if none currently exists, what kinds of measures need to be developed?
7. In light of the answers to the above, what is the evidence that shows the degree to which Macau teachers are sufficiently able, equipped and skilled to provide inclusive education?
8. What specific features of the teacher-education curriculum need to be present or introduced to improve teachers' performance in providing inclusive education? Explaining this will help your reader to understand if and why self-efficacy (or other factors) may be important causal variables.

**The sample**

The paper explains that due to the sample size the long version of the questionnaire was unable to be subjected to the same analysis as the shorter version. However, in the second paragraph of the 'participants' section you do not clarify whether the shorter version was actually completed and, if so, whether this was by a different (unidentified) group of participants or by the same group that completed the longer version.

Points that must be addressed to make the article scientifically sound:

1. Was the shorter version of the questionnaire completed?
2. If so, was it administered to the same group as completed the longer version?
3. If the same group completed both the long and short versions of the questionnaire, what was the order and chronological spacing of the two questionnaires and what steps were

taken to measure and control for any 'learned response' from completing the first questionnaire that may have affected the completed second questionnaire?

4. If the same group completed both versions of the questionnaire, what would the effect have been on your data if the order of completing the questionnaires had been reversed and how do you know?
5. How do these issues affect what can be safely concluded from your data?

### **Presented data**

Measuring the extent to which teachers and those preparing to become teachers feel they are equipped to undertake their professional role may be potentially useful but the interpretation and use of such data must be approached with appropriate caution. Teachers' beliefs about their professional efficacy is likely to be of significant interest to teacher educators, to those responsible for their subsequent professional development during practice and to the individual teachers involved, but such information does not provide us with evidence that they are in fact skilled and experienced in any particular way or to what degree, or that they are able to perform at the level of their beliefs, however sincerely these are held. This is not sufficiently recognised or explored in the present version of the article.

When presenting the internal validity and reliability data for the different models much data and analysis is presented. To avoid the reader becoming lost in this material, it would be helpful to expand the discussion to discuss more fully what such data can be understood to mean and which of the data the authors feel it is more important to focus on and why.

An example of the difficulty this may create for the reader appears on p.8, where the paper states that the total variance explained (TVE) in the 4<sup>th</sup> round of analysis was 63% from a two factor solution, but it would have been very helpful to explain to the reader in more detail what this signifies. Whilst it is not uncommon in the social sciences for researchers to consider TVE values of 60% or even less as acceptable (which is not the case in other fields of research, where much higher values are considered essential), it would be worth reminding the reader what this actually means in practice – i.e. that something(s) which explains 63% of the variance tells us that nearly half of the cause of the variance is unknown and may simply be due to chance and that the analysis therefore fails to explain almost 40% of what is going on.

### Points that must be addressed to make the article more scientifically sound:

1. Remove the present ambiguity in the text that surrounds the discussion of the relationship between teachers' beliefs and their actual performance.
2. Include more clarity about which data is most relevant to the discussion at each particular point (e.g. include more examples of what needs to be focussed on in the data) and provide more help to your reader to better identify and understand which data is most meaningful or informative at any given point in the discussion.

### **Context**

In order to contextualise the study more fully and aid meaningful comparison with other research, the paper could usefully have explained more about the rationale underlying the selection of participants, the criteria used to select them and the demographic data thought to be important to measure. We are told the participant group was comprised of 36 males and 84 females, with 50 being below 25 years of age and 70 being older and of these 44 were preparing to teach in

kindergarten, 58 in primary schools and 18 in secondary schools. The rationale for why these data may have been important and of which population the participants were thought to be representative is not discussed. Sample size is mentioned with regard to factor analysis but not discussed with regard to the population of which the sample is thought to be representative.

Points that must be addressed to make the article more scientifically sound:

1. How was the sample identified and its members selected?
2. What were the inclusion criteria and exogenous variables and how were these selected and analysed to illustrate?
3. Was this designed to be an asymmetric sample, a cross-sectional sample, or some other kind of sample and what was the rationale for the decision?
4. How might the choice of sample have influenced your findings?
5. The sample itself is quite small and the paper helpfully explains that this significantly constrained the conclusions that could be drawn, but it would have been helpful to clarify what conclusions it would have been possible to draw from an appropriately larger sample.

**Discussion**

I would also have welcomed more discussion of the fascinating paragraph starting at the bottom of p.10 which talks about the two- and nine- factor solutions. It would have been interesting to know what the authors feel may be the merits of these solutions in future research on self-efficacy and under what circumstances they might be usefully employed in place of the 12- or 24-item ones.

The cautionary tone adopted in the final paragraph of the discussion section might have been more usefully used to open the discussion, to provide opportunities to refer back to the points made when exploring what the data may be interpreted to be revealing.

The discussion claims evidence exists to support the statement that teacher self-efficacy determines to a strong degree the enabling of effective inclusive education, citing Holzberger *et al.* (2013) and Künsting *et al.* (2016). However, I feel that the authors have confused or somewhat misrepresented the stances taken by these sources, as they do not talk of 'determination' (with its implication of causality) but refer to variables that at best can only 'partially confirm' (Holzberger) or 'predict' or 'mediate' (Künsting) other variables. The Holzberger and Künsting articles are careful not to attribute causality to these and Holzberger is also reminds us of the common misunderstanding of the relationships between causes and consequences. The present paper is less clear on these matters and through its language (such as 'determines') appears to have incorrectly assumed that correlation establishes both the existence and direction of causality.

In the abstract the article says that "*Issues are raised regarding generalizability of scales and the impact of culture, demographics, and edifying issues that may impact on the usefulness of such scales.*" With the exception of generalisability, these topics are not sufficiently or directly explored.

Points that must be addressed to make the article scientifically sound:

1. Adjust the use of language throughout the article to ensure that no reader can conclude that a causal relationship between self-efficacy beliefs and any particular educational



outcome has been established, or is capable of being established, by any of the instruments mentioned.

2. More clarity and accuracy in the use of language is also likely to have an effect on the conclusions that can be and have been drawn in some cases and I encourage the authors to amend any text that is vague in this regard, to avoid a reader concluding that (for example) causality is established or implied.
3. Include a discussion of the impact on self-efficacy and inclusive education by factors such as culture, demographics, and edifying issues that may impact on the reliability of self-report data and instruments seeking to identify self-efficacy.
4. As part of the discussion, explain the contribution the current study is able to make to establishing the impact of culture, demographics, and other edifying issues on teachers' professional practice.

#### **Additional suggestions**

Additional citations would be helpful in places where claims are made but not evidenced, for example when the article says (p.3):

*"As a result of education reform, many schools have established more comprehensive mechanisms for education quality assurance and school effectiveness, and inclusive education, although this is still in its infancy."*

The text would benefit from more careful proof reading to remove occasional misspellings; for example, in the phrase "several initiates have taken place to lead schools toward more fully inclusive practices" (p.3) it seems likely that 'initiates' was meant to be 'initiatives'.

#### Points that must be addressed to make the article scientifically sound:

1. Please check the text for opportunities to enhance the article by supplying supporting citations where statements are made without providing these.
2. Proof read the text carefully to remove misspellings and to improve grammar.

#### **Is the work clearly and accurately presented and does it cite the current literature?**

Partly

#### **Is the study design appropriate and is the work technically sound?**

Partly

#### **Are sufficient details of methods and analysis provided to allow replication by others?**

Partly

#### **If applicable, is the statistical analysis and its interpretation appropriate?**

Partly

#### **Are all the source data underlying the results available to ensure full reproducibility?**

No

**Are the conclusions drawn adequately supported by the results?**

Partly

**Is the argument information presented in such a way that it can be understood by a non-academic audience?**

Partly

**Does the piece present solutions to actual real world challenges?**

Partly

**Is real-world evidence provided to support any conclusions made?**

Partly

**Could any solutions being offered be effectively implemented in practice?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** All areas of educational research, with a special focus on research methods, assessment and the reliability and validity of instruments designed to measure empirical and affective variables using statistical analysis.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 30 November 2020

<https://doi.org/10.21956/emeraldopenres.14612.r27214>

© 2020 Tsigilis N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Nikolaos Tsigilis**

Aristotle University of Thessaloniki, Thessaloniki, Greece

The current manuscript attempts to examine the psychometric properties of a short version (12 items) of the long version of TSES (24 items). Towards this end and in line with prior studies a new translation of the TSES long version in Chinese was conducted. Authors employed various advanced statistical techniques comparing the two versions. My two major concerns about this paper is the small sample size and the resulting side-effects and the reliance only in statistical criteria for excluding items.

Authors correctly state that the size of 120 participants is satisfactory for conducting EFA, because the ratio of the number of participants to the number of variables is around 10:1. While this is true

for EFA, the situation is different for CFA. In particular the ratio is referred to the number of participants to the number of estimated parameters<sup>1</sup>. For a one-factor model with 12 observed variables 36 parameters are being estimated, resulting in a ratio of 3.33. This ratio is even smaller for two or three factor models. Such ratio clearly cannot provide stable estimates.

Moreover, the small sample size prevented the examination of the factorial validity of the long version. Given that the TSES-LV serves as criterion for testing the TSES-SV, and since there are no results for its factorial structure any comparison with the TSES-SV is of limited usefulness. In other words it should first understand the long version before attempting to test a shorter one. As the authors admit this cannot be done due with the current sample size.

Another side-effect of the restricted sample size is the fact that several analyses were conducted and many alternative models were tested using the same data set. This approach entails the threat of capitalization on chance.

Findings showed that a two-factor solution comprising 9 items had adequate fit to the data. However, no explanation is provided (apart of course of statistical criteria) as to why the excluded items did not operate as originally designed. Further elaboration might help author justify their decisions.

## References

1. Rex B Kline: Principles and Practice of Structural Equation Modeling, Fourth Edition. 2015.

### **Is the work clearly and accurately presented and does it cite the current literature?**

Partly

### **Is the study design appropriate and is the work technically sound?**

No

### **Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

### **If applicable, is the statistical analysis and its interpretation appropriate?**

No

### **Are all the source data underlying the results available to ensure full reproducibility?**

No

### **Are the conclusions drawn adequately supported by the results?**

Yes

### **Is the argument information presented in such a way that it can be understood by a non-academic audience?**

Yes

### **Does the piece present solutions to actual real world challenges?**

Yes

**Is real-world evidence provided to support any conclusions made?**

Yes

**Could any solutions being offered be effectively implemented in practice?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Research methods, statistical analyses, psychometrics, education

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

---

## Comments on this article

### Version 1

Reader Comment 11 Jun 2020

**Tim Loreman**, Concordia University of Edmonton, Edmonton, Alberta, Canada

Excellent article yet again from Chris Forlin with colleague Elise Montiero. The data is used well and as always the writing is superb. Forlin's many years of expertise researching topics connected to this one shine through.

**Competing Interests:** I am a friend and long-time research partner of Chris Forlin.

---