

Adaptation of selected self-efficacy scales

Martin Fico

Masaryk University

Abstract

In the presented systematic review, I focus on adapted versions of two widely used self-efficacy scales in the world: the Norwegian Teachers Self-Efficacy Scale (NTSES) and the Teacher Self-Efficacy Scale (TSES). Based on a thorough search in the Web of Science (WOS), I will attempt to answer the following research question: What adaptation and validation procedures did the authors of studies using NTSES and TSES instruments choose, and what were their outcomes? The aim of the article is to help other researchers who are contemplating which scale to adapt, how to proceed with adaptation, and how to verify its psychometric properties. The article covers a total of 16 studies from journals and conference proceedings indexed in Web of Science. Based on the results, it is essential to consider cultural adaptation of the scales and not solely focus on linguistic translation.

Nella presente revisione sistematica, mi concentro sulle versioni adattate di due scale di autoefficacia ampiamente utilizzate nel mondo: la Norwegian Teachers Self-Efficacy Scale (NTSES) e la Teacher Self-Efficacy Scale (TSES). Sulla base di un'accurata ricerca nel Web of Science (WOS), tenterò di rispondere alle seguenti domande di ricerca: Quali procedure di adattamento e convalida hanno scelto gli autori degli studi che utilizzano gli strumenti NTSES e TSES, e quali sono stati i loro risultati? L'obiettivo dell'articolo è aiutare altri ricercatori che stanno valutando quale scala adattare, come procedere con l'adattamento e come verificarne le proprietà psicometriche. L'articolo copre un totale di 16 studi da riviste e atti di conferenze indicizzati nel Web of Science. Dai risultati emerge che è fondamentale considerare l'adattamento culturale delle scale e non concentrarsi solo sulla traduzione linguistica.

Keywords: self-efficacy scale; NTSES; TSES; adaptation of research instruments; cultural adaptation

Parole chiave: self-efficacy scale; NTSES; TSES; adattamento degli strumenti di ricerca; adattamento culturale

1. Introduction

Authors have been addressing teacher self-efficacy for four decades (Berg & Smith, 2018), and a considerable number of research tools designed for these purposes are available. However, for a researcher, it is crucial that their measurement is reliable, the tool exhibits acceptable validity, and the results can be meaningfully compared with those of other authors from their own or foreign countries. This can also contribute to the replication of research findings, a point of critique in the context of psychological research (Yong, 2012). Perhaps due to these reasons, some researchers opt for a previously validated research scale and attempt to adapt it to conditions specific to their target group and country. Adapting foreign research instruments is often a challenging process with uncertain outcomes, making selecting the appropriate scale pivotal.

But how does one decide amidst this multitude of instruments? The chosen tool must encompass the studied phenomenon in its utmost complexity, exhibit suitable psychometric properties, and ideally have been cross-culturally adapted. In this article, I build upon the works of Smetáčková et al. (2017) and Berg & Smith (2018), who identify two relevant and internationally successful measurement tools for teacher self-efficacy: Norwegian Teachers Self-Efficacy Scale (NTSES) and Teacher Self-Efficacy Scale (TSES). Within this systematic review, my objective is to assist fellow researchers in deciding which tool to adapt. I will concisely introduce both instruments to readers and elucidate their adapted foreign forms along with the associated evidence.

2. Theoretical background

2.1 Teacher self-efficacy

Teacher self-efficacy derives from Bandura's (1997) socio-cognitive theory, which discusses the subjective belief in attaining a certain level of performance. Despite being a subjective belief rather than actual performance, numerous researchers demonstrate a link between self-efficacy and real professional performance (Goddard et al., 2004; Karim et al., 2021). The description of how this connection between self-efficacy and performance functions is outlined by Lippke (2020). She explains that the sources of self-efficacy, our subjective conviction, encompass our own experiences, vicarious experiences, verbal persuasion, and emotional arousal. Based on how these sources shape our self-efficacy, they manifest in our actions, such as in choice (approach, avoid), effort and persistence, thinking and decision-making, and emotional reactions (stress, depression, etc.) (Lippke, 2020).

Teacher self-efficacy is directed towards various aspects of the teaching profession, often grounded in the pedagogical or pedagogical-psychological domain. Some authors have endeavored to construct more comprehensive instruments aimed at capturing teacher self-efficacy across a broader spectrum of their profession (Gibson & Dembo, 1984; Skaalvik & Skaalvik, 2007; Tschannen-Moran & Woolfolk Hoy, 2001). In other studies, we encounter scales that measure only one or fewer areas of teacher self-efficacy (e.g., Fico, 2021). In the past, the originator of the self-efficacy concept, Bandura himself, devoted attention to teacher self-efficacy and even proposed a research instrument. However, it failed to gain recognition, perhaps due to its less clear theoretical, pedagogical background or undisclosed psychometric properties (Smetáčková et al., 2017).

2.2 Adaptation of research instruments

The concept of adapting a research instrument entails linguistic and cultural adjustments of items (and the entire instrument) within the context where the scale will be employed. Several internationally recognized guidelines for the process of linguistic and cultural adaptation exist (Beaton et al., 2000; Gudmundsson, 2009; Sousa & Rojjanasrirat, 2010; Hernandez et al., 2020), supplemented by new insights from other authors' studies (Epstein et al., 2015; Gana et al., 2020). The adaptation process may differ depending on the author, who chooses

either a predefined guideline approach or a different one. Thorough and diligently conducted adaptation can lead to replication of research results in different contexts, a concern that Yong (2012) raised in psychological research, including self-efficacy research.

Linguistic adaptation involves expert translation from the original language to the target language. Typically, this involves multiple independent translations, usually at least two, followed by synthesis (e.g., Delgado-Lobete et al., 2021). Subsequently, the translated scale can be assessed by experts (Prosen et al., 2021) or by the authors of the original scale (Borsa et al., 2012) to ensure that the items truly correspond to the original meaning. An additional recommended step is back-translation into the original language, followed by a comparison with the original scale (e.g., Kara et al., 2006). Deficiencies in the translated scale are then rectified, and the process of assessment and back-translation is reiterated. Cultural adaptation involves not only linguistically accurate translation but also content adaptation of items to match the country from which the research sample originates (Gudmundsson, 2009; Schendel & Tolmie, 2016; Epstein et al., 2015).

3. Measurement of teacher self-efficacy (NTSES and TSES)

3.1 NTSES

The Norwegian Teacher Self-Efficacy Scale (NTSES) was developed by Skaalvik & Skaalvik (2007) for the Norwegian context and environment. According to Berg & Smith (2018), NTSES is a tool that aligns more closely and rigorously with Bandura's (1997) theory compared to other instruments. When creating the 24-item scale, Skaalvik & Skaalvik (2007) followed Bandura's (2006) recommendation that self-efficacy measures perceived capability, not intention. This is evident in the phrasing of NTSES items, which focus on "can do" rather than "will do", as highlighted by Berg & Smith (2018). NTSES consists of six dimensions, each containing four items, and the original version is in Norwegian. Skaalvik & Skaalvik (2007) verified the internal consistency of NTSES using both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), confirming the appropriateness of the 6-dimensional model with four items each. These six dimensions make NTSES a potentially valuable instrument for measuring self-efficacy, as it captures the teaching profession within a broader context and provides more precise insights into teachers' self-efficacy. The mentioned dimensions are: Instruction; Adapting education to individual students' needs; Motivating students; Keeping discipline; Cooperating with colleagues and parents; and Coping with changes and challenges.

3.2 TSES/OSTES

The Teachers' Sense of Efficacy Scale (TSES), also known as the Ohio State Teachers Efficacy Scale (OSTES), was developed by Tschannen-Moran et al. (1998) and was regarded by some researchers (Duffin et al. 2012) as the best tool for measuring teacher self-efficacy until that point. The authors (Tschannen-Moran et al., 2001) transparently describe the scale's development process and the changes they had to implement due to fluctuating results in factor structure. TSES is currently used in two versions: a short 12-item version and a long 24-item version (Berg & Smith, 2018). TSES is among the most widely used instruments for measuring teacher self-efficacy and is employed in many languages and countries, including research within TALIS. Both the 12-item and 24-item versions consist of three subscales, each containing up to 8 items. These subscales are: Efficacy in student engagement; Efficacy in instructional practices; and Efficacy in Classroom Management.

4. Systematic review¹

4.1 Objectives of the overview study and selection of empirical studies

Martin Fico – *Adaptation of selected self-efficacy scales*

DOI: <https://doi.org/10.6092/issn.1970-2221/19242>

The aim of the presented text is to elucidate the adapted versions of two globally employed research instruments measuring teacher self-efficacy. Addressing the research questions in this overview study can assist fellow researchers in deciding which research tool to embrace, how to adapt it, and subsequently employ it in the study of teacher self-efficacy.

I approach this overview study with the following research questions:

- What adaptation and validation procedures did authors of studies employing the NTSES tool choose, and what were their outcomes?
- What adaptation and validation procedures did authors of studies employing the TSES/OSTES tool choose, and what were their outcomes?

Selection of the studies

I curated the research texts for the overview study in the first quarter of the year 2023. I conducted searches in three steps within the Web of Science database, focusing on the English language. In each step, my focus was on finding studies that described the adaptation of a specific tool (NTSES, TSES/ OSTES). The keywords for the search comprised the full tool name, its acronym, and the phrase “teachers’ self-efficacy”. Example keywords are the following: Norwegian teachers self-efficacy scale; NTSES; teachers’ self-efficacy. The selection encompassed articles in journals as well as those in conference proceedings. During broader searches in other databases, I encountered additional texts discussing the adaptation of these specific tools; however, these were not indexed in Web of Science and, therefore, were not included in the selection for the overview study. The final list of studies can be viewed in Table 1.

- NTSES: Upon configuring the search, six results emerged. One result was immediately excluded, as it pertained to a research tool with a different focus, rendering it irrelevant to this work. After reading the abstracts, five remaining studies were deemed relevant and were included in the narrower selection. One text was written in Polish. Consequently, five texts discussing adapted versions of the NTSES tool were incorporated into the overview study.
- OSTES: The research tool TSES is sometimes denoted as OSTES as well; thus, I separated it in the search and conducted a distinct search. Four results emerged for OSTES, yet three of them did not discuss adaptation or work with an adapted OSTES tool. Therefore, only one study aligned with the objectives of the work and was included in the selection.
- TSES: The popularity of the TSES tool became evident upon entering the keywords, with a staggering 76 results, surpassing the other tools by far. Among these 76 results, 53 were relevant to my search for TSES in the desired context. After reviewing the abstracts, 35 texts were excluded as they did not utilize TSES in different countries or involve its adaptation or modification. The remaining 18 results were subjected to further reading, of these 10 aligned with the objectives of the overview study, encompassing various linguistic and cultural adaptations across diverse countries. One study was subsequently excluded from the selection, as it solely provided a secondary validation of an already included adapted tool. As a result, the overview study includes ten works.

Table 1. Selected studies for systematic review

Self-Efficacy Scale	Country	Language of scale	Type of text
NTSES	Turkey, Iran	Not in text	Proceedings paper

NTSES	Serbia	Not in text	Proceedings paper
NTSES	Czech Republic	Czech	Proceedings paper
NTSES	Italy	Italian	Article
NTSES	Poland	Polish	Article
TSES (OSTES)	Slovakia	Slovak	Article
TSES	Serbia	Serbian	Article
TSES	France	French	Article
TSES	China	Chinese	Article
TSES	Mexico	Spanish	Article
TSES	China, Japan, Korea	Chinese, Japanese, Koreanese	Article
TSES	Spain	Spanish	Article
TSES	Greece	Greek	Article
TSES	Nigeria	French	Article
TSES	Malaysia	Malaysian	Article
TSES	Poland	Polish	Article

5. Results

5.1 NTSES

Adaptation procedure

The most thoroughly detailed adaptation procedure for NTSES was presented by Avanzi et al. (2013) and Baka (2017). In both studies, independent translators translated the original tool from English, involving two translators for Baka (2017) and three for Avanzi et al. (2013). Subsequently, items showing differences were modified. After creating a unified translated version, they consulted practicing teachers (Avanzi et al. 2013) to ensure the best understanding of items in the field and made minor adjustments. A back-translation followed (Avanzi et al., 2013; Baka, 2017) to ensure the meaning of individual items in the original NTSES was retained. Baka (2017) piloted the final scale on a sample of 30 teacher education students. Avanzi et al., (2013) ensured a reliable translation with corresponding content by consulting the authors of the original NTSES (Skaalvik & Skaalvik, 2007).

Other researchers were less transparent in describing the adaptation process or did not describe it at all. The authors of the Czech translation of NTSES merely state that they translated items into Czech without specifying whether from the original Norwegian scale or its English version (Dofková & Kvintová, 2017). The authors from Serbia provide a modest description of the adaptation process, stating only that they used the NTSES tool,

without mentioning the languages involved or the process itself (Djigić et al., 2014). Khezerlou (2013) states that they shortened the scale to 14 items without detailing the process or language involved. Interestingly, Khezerlou (2013) conducted the research in both Turkey and Iran, potentially implying two adaptations (at least linguistically), although no information about this is present in the article.

Research sample

The sizes of the research samples and the target groups varied across studies employing the adapted NTSES. The smallest sample consisted of only 49 teacher education students (Dofková & Kvintová, 2017), while the largest comprised 404 teachers (Baka 2017). Practicing teachers constituted the research sample in the remaining three studies, involving 168 teachers in Serbia (Djigić et al., 2014), 156 and 230 in Turkey and Iran (Khezerlou, 2013), and 347 in the case of Avanzi et al., (2013). The selection of respondents was based on availability (Dofková & Kvintová, 2017; Djigić et al., 2014; Khezerlou, 2013; Avanzi et al. 2013; Baka 2017). The authors did not aim to approximate the distribution of the general population. Avanzi et al. (2013) specified the respondent selection and engagement process more comprehensively, narrowing the selection to four specific schools in one city and involving school principals to approach teachers. None of the studies mention the response rate of questionnaires.

Validation

The original NTSES (Skaalvik & Skaalvik, 2007) is a six-dimensional tool, with its internal consistency validated through Exploratory Factor Analysis (EFA) and subsequently Confirmatory Factor Analysis (CFA). Two studies (Avanzi et al., 2013; Baka 2017) also opted for CFA to validate the six-dimensional model in their adapted versions. The evidential basis for their calculations and findings included Comparative Fit Index (CFI), Non-Normed Fit Index (NNFI), Root Mean Square Error of Approximation (RMSEA), and χ^2/df (Avanzi et al., 2013), as well as RMSEA, Goodness of Fit Index (GFI), and Adjusted Goodness of Fit Index (AGFI) (Baka 2017). Reliability was assessed by authors, and Cronbach's alpha was declared for each dimension (Avanzi et al., 2013; Baka 2017). Avanzi et al., (2013) also examined test-retest reliability, composite reliability (CR), and average variance extracted (AVE). In other studies, validation either did not occur or was not published (Dofková & Kvintová, 2017; Djigić et al., 2014; Khezerlou, 2013). Khezerlou (2013) reported the reliability of the entire tool through Cronbach's alpha, while Djigić et al. (2014) declared reliability for both the overall tool and individual dimensions.

Psychometric results

Based on good fit indexes (CFI=0.98; NNFI=0.98; RMSEA=0.57; $\chi^2/df=2.96$) and high reliability values ($\alpha \geq 0.8$; $CR \geq 0.8$; $AVE \geq 0.52$), Avanzi et al. (2013) stated that the adapted tool exhibited favorable properties and a consistent six-dimensional internal structure. They suggested the adapted tool is suitable for international comparisons of teacher self-efficacy results. The 6-factor model corresponds to the original model. Based on CFA results, Baka (2017) rejected the 6-factor structure and found the 3-factor model appropriate. This 3-factor model included factors: motivate pupils and adapt instructions to individual needs, maintain discipline and cooperate with parents, instruction and cooperate with colleagues. Baka supported this 3-factor model with CFA evidence (GFI=.917; AGFI=.898; RMSEA=.073). In the remaining studies (Dofková & Kvintová, 2017; Djigić et al., 2014; Khezerlou, 2013), psychometric properties of the adapted NTSES were not discussed, but they worked with it as a six-dimensional tool.

5.2 TSES

Adaptation procedure

For the Slovak version of OSTES, authors (Novocký & Rovňanová, 2021) referred to Gavora (2011) for the adaptation process. Gavora had already conducted the cultural and linguistic adaptation of OSTES for use in Slovakia. Gavora's adaptation involved not only translation into Slovak but also adjusting items to match the Slovak context and enhance comprehension for the target group. Novocký & Rovňanová (2021) subsequently made further adjustments to the adapted OSTES version. They reformulated some items to suit better their target group, which differed from Gavora's (2011) target group. They removed certain items and changed the Likert scale from a nine-point scale to a five-point scale, justified by the research purpose. Other authors utilized independent translators for translating from the original, with varying numbers of translators, from an unspecified number (Jang, Cho & Wiens 2019) to five independent translators (Koniewski, 2019). All authors used back-translation, and some authors proceeded to make changes to specific items or remove them (Koniewski, 2019; Valls et al., 2020; Tsigilis et al., 2010; Jang et al., 2019; Salas-Rodriguez et al., 2021; Ninkovic & Floric, 2018; Khairani & Makara, 2020; Ruan et al., 2015). Expert assessment was present in some studies; for example, Valls et al. (2020) consulted a self-efficacy expert, Salas-Rodriguez et al., (2021) consulted a language expert, Ruan et al. (2015) consulted language and empirical research experts, and Burgueno et al. (2018) consulted experts in teacher education.

Research sample

Various approaches were taken for selecting research participants. Valls, Bovin & Benoit (2020), Tsigilis et al. (2010), and Koniewski (2019) assembled their samples from primary and secondary teachers. Novocký & Rovňanová (2021), Jang et al. (2019), Salas-Rodriguez, Lara & Martinez (2021), Ruan et al. (2015), and Pintus et al. (2021) did not specify the teaching level but worked with in-service teachers. Other authors worked with combinations of teachers from multiple levels (Ninkovic & Floric 2018), teachers and students (Khairami & Makara, 2020), only student teachers (Burgueno et al., 2018), or special education teachers (Lu et al., 2020). The sample size was another significant consideration. The number of respondents often depended on the statistical operations researchers intended to perform with the data. The necessary sample size for factor analyses is typically recommended based on the number of items in the research scale. The size of the research samples for TSES adaptations ranged from 190 (Salas-Rodriguez et al., 2021) to 4465 (Koniewski, 2019). None of the mentioned studies utilized a purely representative sample.

Validation

For the statistical validation of the adapted research tool, researchers frequently recommend using Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), or their combination (Orcan et al., 2018). Authors of all the mentioned studies chose one of these options. EFA was used by Pintus et al. (2021), and Novocký & Rovňanová (2021) employed Principal Component Analysis (PCA), while others favored CFA. Authors using EFA/PCA supported their findings with factor loadings, inter-item correlations, Kaiser-Meyer-Olkin (KMO) Test, Bartlett's test, Eigenvalues, and Variance percentages. Authors employing CFA supported their results with the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Standardized Root Mean Square Residual (SRMR), and RMSEA. When comparing two or more models, some studies also reported Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). However, not all commonly used values

that indicate model accuracy and internal consistency were documented in some studies; for example, TLI was not published in certain studies (Tsigilis et al., 2010; Ninkovic & Floric, 2018; Jang et al., 2019; Koniewski, 2019).

While describing statistical methods, the authors provided important cut-off points for individual indicators with references to other authors. However, the interpretation of these values and the strictness of assessment varied. Lu et al. (2020) and Rvan et al. (2015) worked with stricter criteria, while others referred to lower thresholds or even greater flexibility in their results.

Results

Valls et al. (2020) validated with CFA both a 1-factor and a 3-factor model of the 12-item TSES for the French context. While the 1-factor model was not confirmed in CFA, the 3-factor model proved to be internally consistent, and its factor structure corresponds to the original TSES (Tschannen-Moran & Woolfolk Hoy, 2001). Detailed CFA results are displayed in Table 2.

Table 2. CFA results Valls et al. (2020)

Model	GFI	CFI	TLI	ECVI	AIC	RMSEA	SRMR	x2	df	x2/df
1-f	.75	.69	.63	1.76	496.56	.161	.102	448.56	54	8.31
3-f	.90	.90	.87	.85	239.28	.097	.061	185.28	51	3.63

Similar results were also recorded by Ninkovic & Floric (2018), who adapted and tested a short form of the scale, i.e., a 12-item one. CFA confirmed the 3-dimensional model, but the authors had to make several adjustments in individual items. Detailed CFA results is displayed in Table 3.

Table 3. CFA results of Ninkovic & Floric (2018)

Model	x2	p	df	CFI	RMSEA	SRMR	AIC	BIC
1-f	384.462	<.001	54	.827	.135	.069	16454.638	16602.731
3-f	191.371	<.001	51	.927	.090	.045	16198.153	16358.586
mod 3-f	142.134	<.001	50	.953	.074	.040	16132.634	16297.181

The 12-item TSES scale was also adapted and tested by Rodriguez et al. (2021). The 1-factor solution was not confirmed in the CFA, but the 3-factor solution with borderline results was (RMSEA value was higher than recommended thresholds and TLI just below the cutoff score of 0.90). Detailed CFA results are displayed in Table 4.

Table 4. CFA results of Rodriguez et al. (2021)

Model	x2/df	TLI	CFI	SRMR	RMSEA
3-f	2.97	.886	.912	.053	.102



Major adjustments had to be made by Khairani & Makara (2020), who failed to achieve even one of the originally intended structures. Neither the 3-factor nor the 1-factor version of the TSES corresponding to the original instrument has been validated. After translating the three items into a factor other than the original factor, they declared a usable 2-factor and 3-factor modified structure. These claims were substantiated by the results from the CFA, while the minimum TLI was below the declared cut-off point in both cases. The findings were summarized by the authors themselves: “Moreover, there are many elements of reliability and validity that should be considered before having strong confidence in a questionnaire”. (Khairani & Makara 2020, p. 321). Detailed CFA results are displayed in Table 5.

Table 5. CFA results of Khairani & Makara (2020)

Model	χ^2/df	p	CFI	TLI	RMSEA	SRMR
Malaysia 3-f	2.35	< 0.001	.883	.870	.066	.054
Malaysia 2-f	2.22	< 0.001	.910	.895	.084	.054

Similar CFA results were also obtained by Koniewski (2019), for whom some items were loaded on different factors than originally intended. CFI results are below the cut-off point, and SRMR and RMSEA are at an acceptable level. Based on the results, Koniewski (2019) reflects on the need to modify the scale when using it in the following measurement and another research.

The 2-factor model was also published by Novocký & Rovňanová (2021), who revealed it by PCA. They justified the choice of this method by the changes they made in the tool. Based on factor loadings (< 0.30), they eliminated 3 items from the originally adapted instrument (Gavora, 2012). Other items achieved a factor loading of at least 0.4. The results of KMO (0.891) and Bartlett test (0.000) confirmed the adequacy of PCA. The results of the analysis showed that a two-dimensional solution is suitable. Efficacy in using teaching strategies with eight items and Efficacy in class management with four items.

Authors from Nigeria (Jang et al., 2019) also reduced the number of items who had to drop seven items from the original 24-item instrument based on the CFA results. After their removal, the results of the CFI, SRMR, and RMSEA indicators showed the suitability of the 3-factor model, although they did not support, for example, the results of the TLI indicator. Such a significant change in the tool is justified by the need for adaptation not only linguistically but also culturally (Jang et al., 2019). Detailed CFA results are displayed in Table 6.

Table 6. CFA results of Jang et al. (2019)

Model	χ^2	df	SRMR	RMSEA	CFI
Niger 3-f	272.246	111	.041	.048	.937

Researchers from Italy also emphasize cultural adaptation (Pintus et al., 2021). Based on EFA, they preferred the 4-factor model over the original 3 and 1 factor versions. 2 factors correspond to the original 3-factor TSES (Tschannen-Moran & Woolfolk Hoy, 2001) – “Classroom Management” and “Student Engagement”. The other 2 factors were named “Inclusive instructional strategies learner-centered” and “Traditional instructional

strategies teacher-centered”. They discuss this extension of the instrument with the NTSES instrument, which contains six factors.

In the Spanish version of the TSES, Burgunéo et al. (2018) attempted to adapt and validate both the 12-item model and the 24-item model. CFA did not confirm the original factor structure of the 24-item model, and the authors evaluated the 12-item model as more appropriate. However, CFA indicated that the most suitable Spanish TSES is a modified 11-item and 3-factor model. Detailed CFA results are displayed in Table 7.

Table 7. CFA results of Burgunéo et al. (2018)

Model	x2	df	x2/df	CFI	IFI	TLI	SRMR	RMSEA	AIC
Spain 3-f	69.6	41	1.70	.979	.980	.972	.031	.052	119.6

Two independent research teams from Asia validated the 12-item and 24-item TSES (Ruan et al., 2015; Lu et al., 2020). Both teams used CFA for validation and adhered to strict cut-off points. Neither research accepted the results of language adaptation in either model, so he had to proceed with a more radical adjustment. They created an 11-item, 3-factor model called the Asian short form. The model constructed in this way emphasized not only linguistic adaptation but also cultural adaptation and achieved excellent values in the CFI, TLI, RMSEA, and SRMR indicators. Detailed CFA results are displayed in Table 8 and Table 9.

Table 8. CFA results of Ruan et al. (2015)

Model	x2	df	CFI	TLI	RMSEA	SRMR	WRMR
Asian 3-f	78.715	41	.975	.966	.047	.033	.871

Table 9. CFA results of Lu et al. (2020)

Asian model	x2	df	n	p	TLI	CFI	RMSEA	SRMR
China 3-f	62.457	41	133	.017	.951	.964	.063	.049
Korea 3-f	55.078	41	198	.070	.980	.985	.042	.037
Japan 3-f	41.128	41	147	.465	1.000	1.000	.005	.029

The only author from the mentioned studies who managed to adapt the 24-item, 3-factor TSES instrument was Koniewski (2019).

6. Discussion

The approach authors select for adapting instruments in their studies is undoubtedly chosen with good intentions – aiming to effectively adapt a validated foreign instrument for use in a new context. In certain selected texts, the question arises whether this constitutes adaptation or merely a thorough linguistic translation. Gana et al. (2021) also highlight this issue. When examining individual studies from the process of adaptation to the psychometric properties of the instrument, it becomes evident that the best results are achieved by authors who

not only adjusted the instrument linguistically but also culturally (through item changes or reductions) and transparently describe this in their studies (e.g., Ruan et al., 2015; Lu et al., 2020; Burgunéo et al., 2018). Multiple independent translations, an internationally recognized and recommended practice (Orcan, 2018; Gudmundsson, 2009; Hernandez et al., 2020), can be seen in all selected studies describing translation procedures. Authors still employ back-translation, although its necessity is questioned in certain texts (Gana et al., 2021; Epstein et al., 2015). In contrast to back-translation, there is the notion that expert committee assessment is advisable (Epstein et al., 2015), a practice often overlooked or omitted in texts.

Gana et al. (2021) express a belief that the era has ended where one study with an available student sample is enough to validate the adapted research tool. The composition of research samples in the selected studies does not reflect this – often relying on student participants. On the other hand, every researcher conducting quantitative research likely understands the challenge of assembling a good research sample. Sample sizes often fall short from a statistical recommendation perspective or, at the very least, raise discussion in relation to selected statistical operations. In literature, various recommendations for EFA exist, such as 10-15 respondents per item (Soukup & Kočvarová, 2016) 100-250 respondents (Cattell, 1978), 300 or more respondents (Field 2009), or even 500 or more (Comrey & Lee, 1992). For CFA, the usual guideline is 20 respondents per item (Schumacker & Lomax, 2010; Kline, 2016). In the selected studies in this text, we also encounter considerably smaller research samples, a stance supported by some in the literature (e.g., Sapnas & Zeller, 2002), and this can also be observed in studies from other scientific fields (e.g., Kara et al., 2006). The issue of research samples surely applies to studies focused on tool validation and is something to consider.

The verification of psychometric properties of adapted questionnaire versions is recommended through CFA (Gana et al., 2021) or both EFA and CFA (Borsa et al., 2012). Many authors from the selected studies adhered to these recommendations, although this wasn't the case in some instances (Dofková & Kvintová, 2017; Djigić et al., 2014; Khezerlou, 2013). Studies occasionally confuse PCA with EFA, as highlighted by Rabušić et al. (2019), which can also be observed in Novocký & Rovňanová (2021), for instance. However, Fabrigar et al. (1999) argue that this is incorrect and, in such cases, EFA should be exclusively used. Nevertheless, studies focused on questionnaire adaptation occasionally showcase alternative approaches, such as Araujo et al. (2010), who employed PCC and ICC, though this wasn't present among the authors of the selected texts. Finally, readers should note the variety of published indicators from EFA and, primarily, CFA. When some studies lack certain crucial indicators, evidence about successful or unsuccessful adaptation is insufficient. This is important to observe and consider when deciding whether to adopt a given tool and which procedure appears successful based on previous studies.

7. Conclusion

Both research instruments for measuring teacher self-efficacy appear suitable for international result comparisons and, therefore for their international adaptation. Repeated adaptations of the TSES instrument worldwide likely make it an increasingly attractive choice for adaptation due to its more straightforward factor structure (1-factor and 3-factor) and broader applicability (short and long versions). However, the successful and transparent adaptation of NTSES in Italy demonstrates an intriguing option for those wishing to measure teacher self-efficacy across multiple dimensions. Unsuccessful or conflicting adaptations underscore the importance of considering linguistic and cultural adaptation, ensuring that the instrument is well-aligned with the target group in the specific country. Within this context, it can be effective to move beyond strict back-translation approaches and instead modify question wordings, replace items, or remove them. In conclusion, it's important to emphasize that transparently disclosing the adaptation process and sharing as many results from statistical

analyses as possible are crucial aspects during the adaptation of an instrument. This ensures that future authors can make informed decisions about whether and how to adapt the given tool. As we can see from the studies presented, it cannot be definitively said which of the mentioned research instruments is more suitable for international adaptation and use in a new context. Both tools appear to be adaptable, but it depends on the adaptation process and subsequent validation. However, for better reproduction of educational or psychological research, it is crucial that authors transparently disclose all procedures and indicators related to these processes. Even an unsuccessful adaptation or validation can serve as advice for other authors, informing them that they need to seek a different, more suitable tool.

Acknowledgments

This paper is supported by an internal grant of Masaryk University, through a project of Specific research (MUNI/A/1496/2023).

Note

1. Used shorts: EFA - Exploratory Factor Analysis; PCA - Principal Component Analysis; CFA - Confirmatory Factor Analysis; PCC - Pearson's Correlation Coefficient; ICC - Intraclass Correlation Coefficient.

Bibliografia

- Araújo, L. G., Lima, D. M. F., Sampaio, R. F., & Pereira, L. S. M. (2010). Escala de Locus de controle da dor: adaptação e confiabilidade para idosos. *Revista Brasileira De Fisioterapia*, 14(5), 438–445. <https://doi.org/10.1590/s1413-35552010000500014>
- Avanzi, L., Miglioretti, M., Velasco, V., Balducci, C., Vecchio, L., Fraccaroli, F., & Skaalvik, E. M. (2013). Cross-validation of the Norwegian Teacher's Self-Efficacy Scale (NTSES). *Teaching and Teacher Education*, 31, 69–78. <https://doi.org/10.1016/j.tate.2013.01.002>
- Baka, Ł. (2017). Norwegian Teacher Self-Efficacy Scale – Psychometric properties of the Polish version of the scale. *Medycyna Pracy*. <https://doi.org/10.13075/mp.5893.00569>
- Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of Cross-Cultural adaptation of Self-Report measures. *Spine*, 25(24), 3186–3191. <https://doi.org/10.1097/00007632-200012150-00014>
- Berg, D. a. G., & Smith, L. F. (2018). The effect of school-based experience on preservice teachers' self-efficacy beliefs. *Issues in Educational Research*, 28(3), <http://www.iier.org.au/iier28/berg.pdf>.
- Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Adaptação e validação de instrumentos psicológicos entre culturas: algumas considerações. *Paidéia*, 22(53), 423–432. <https://doi.org/10.1590/s0103-863x2012000300014>
- Burgueño, R., Sicilia, Á., Casaubón, J. M., Alcaraz-Ibáñez, M., & Lirola, M. (2018). Psychometry of the teacher's sense of efficacy scale in Spanish teachers' education. *Journal of Experimental Education*, 87(1), 89–100. <https://doi.org/10.1080/00220973.2018.1423542>

- Delgado-Lobete, L., Montes-Montes, R., Méndez-Alonso, D., & Prieto-Saborit, J. A. (2021). Cross-Cultural Adaptation and Preliminary Reliability of the Adolescents and Adults Coordination Questionnaire into European Spanish. *International Journal of Environmental Research and Public Health*, 18(12), 6405. <https://doi.org/10.3390/ijerph18126405>
- Djigic, G., Stojiljković, S., & Dosković, M. (2014). Basic personality dimensions and teachers' self-efficacy. *Procedia - Social and Behavioral Sciences*, 112, 593–602. <https://doi.org/10.1016/j.sbspro.2014.01.1206>
- Dofková, R., & Kvintová, J. (2017). Development of the attitude to mathematics in future primary school teachers as a determinant of their self-efficacy. In L. Gómez Chova, A. López Martínez, I. Candel Torres (Eds.). *Edulearn Proceedings 2017*. Iated Academy. <https://doi.org/10.21125/edulearn.2017.1726>
- Duffin, L. C., French, B. F., & Patrick, H. T. (2012). The Teachers' Sense of Efficacy Scale: Confirming the factor structure with beginning pre-service teachers. *Teaching and Teacher Education*, 28(6), 827–834. <https://doi.org/10.1016/j.tate.2012.03.004>
- Epstein, J., Osborne, R. H., Elsworth, G., Beaton, D., & Guillemin, F. (2015). Cross-cultural adaptation of the Health Education Impact Questionnaire: experimental study showed expert committee, not back-translation, added value. *Journal of Clinical Epidemiology*, 68(4), 360–369. <https://doi.org/10.1016/j.jclinepi.2013.07.013>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <https://doi.org/10.1037/1082-989x.4.3.272>
- Gana, K., Boudouda, N. E., Youssef, S. B., Calcagni, N., & Broc, G. (2021). Adaptation transculturelle de tests et échelles de mesure psychologiques : guide pratique basé sur les Recommandations de la Commission Internationale des Tests et les Standards de pratique du testing de l'APA. *Pratiques Psychologiques*, 27(3), 223–240. <https://doi.org/10.1016/j.prps.2021.02.001>
- Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76(4), 569–582. <https://doi.org/10.1037/0022-0663.76.4.569>
- Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2004). Collective Efficacy Beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher*, 33(3), 3–13. <https://doi.org/10.3102/0013189x033003003>
- Gudmundsson, E. H. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, 61(2), 29–45. <https://doi.org/10.1027/1901-2276.61.2.29>
- Hernández, A., Hidalgo, M. D., Hambleton, R. K., & Gómez-Benito, J. (2020). International Test Commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32(3), 390–398. <https://doi.org/10.7334/psicothema2019.306>
- Jang, B. G. (n.d.). Self-Efficacy and quality of classroom interactions of EFL teachers in Niger. *Digital Scholarship@UNLV*. https://digitalscholarship.unlv.edu/tl_fac_articles/273/
- Khairani, A. Z., & Makara, K. (2020). Examining the factor structure of the teachers' sense of efficacy scale with Malaysian samples of in-service and pre-service teachers. *Pertanika Journal of Social Science and Humanities*, 28(1), 309–323.

- Kara, M., Van Der Bijl, J. J., Shortridge-Baggett, L. M., Aşti, T., & Ergüney, S. (2006). Cross-cultural adaptation of the diabetes management self-efficacy scale for patients with type 2 diabetes mellitus: Scale development. *International Journal of Nursing Studies*, 43(5), 611–621. <https://doi.org/10.1016/j.ijnurstu.2005.07.008>
- Karim, M. R., Masud, N. A., Subarna, M. T. N., Billah, M. M., & Wienaar, P. (2021). Self-efficacy: a key components of teacher effectiveness. *Asian Journal of Education and Social Studies*, 25(1) 24–34. <https://doi.org/10.9734/ajess/2021/v25i130590>
- Khezerlou, E. (2013). Teacher self-efficacy as a predictor of job burnout among Iranian and Turkish EFL teachers. *Procedia - Social and Behavioral Sciences*, 70, 1186–1194. <https://doi.org/10.1016/j.sbspro.2013.01.175>
- Lippke, S. (2020). *Self-Efficacy Theory*. Springer eBooks (pp. 4722–4727). https://doi.org/10.1007/978-3-319-24612-3_1167
- Ninković, S., & Knežević-Florić, O. (2018). Validation of the Serbian version of the teachers' sense of efficacy scale (TSES). *Zbornik Instituta Za Pedagoska Istrazivanja*, 50(1), 72–92. <https://doi.org/10.2298/zipi1801072n>
- Novocký, M., & Rovňanová, L. (2021). Relation between the implementation frequency of reflexive methods and self-efficacy of lower secondary school teachers. *Ad Alta*, 11(1), 214–222. <https://doi.org/10.33543/1101214222>
- Orçan, F. (2018). Exploratory and confirmatory factor analysis: Which one to use first? *Journal of Measurement and Evaluation in Education and Psychology*, 9(4), 414–421. <https://doi.org/10.21031/epod.394323>
- Pintus, A., Bertolini, C., Scipione, L., & Antonietti, M. (2021). Validity and reliability of the Italian version of the Teachers' Sense of Efficacy Scale. *International Journal of Educational Management*, 35(6), 1166–1175. <https://doi.org/10.1108/ijem-08-2019-0279>
- Prosen, M., Kvas, A., Bošković, S., & Ličen, S. (2021). Cross-cultural adaptation and psychometric evaluation of the Slovenian version of the nurse professional competence scale. *BMC Nursing*, 20(1). <https://doi.org/10.1186/s12912-021-00664-6>
- Ruan, J., Nie, Y., Hong, J., Monobe, G., Zheng, G., Kambara, H., & You, S. (2015). Cross-Cultural validation of teachers' sense of efficacy scale in three Asian countries. *Journal of Psychoeducational Assessment*, 33(8), 769–779. <https://doi.org/10.1177/0734282915574021>
- Salas-Rodríguez, F., Lara, S., & Martínez, M. (2021). Spanish version of the Teachers' Sense of Efficacy scale: An adaptation and Validation study. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.714145>
- Sapnas, K. G., & Zeller, R. A. (2002). Minimizing sample size when using exploratory factor analysis for measurement. *Journal of Nursing Measurement*, 10(2), 135–154. <https://doi.org/10.1891/jnum.10.2.135.52552>
- Schendel, R., & Tolmie, A. (2016). Beyond translation: adapting a performance-task-based assessment of critical thinking ability for use in Rwanda. *Assessment & Evaluation in Higher Education*, 42(5), 673–689. <https://doi.org/10.1080/02602938.2016.1177484>
- Schumacker, R. E., & Lomax, R. G. (2010). *A Beginner's Guide to Structural equation modeling*, Routledge.
- 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. <https://doi.org/10.1037/0022-0663.99.3.611>

- Smetáčková, I., Topková, P., & Vozková, A. (2017). Development And Piloting Of Teacher Self-Efficacy Scale. *Lifelong Learning*, 7(2), 26–46. <https://doi.org/10.11118/lifele2017070226>
- Soukup, P., & Kočvarová, I. (2016). Velikost a reprezentativita výběrového souboru v kvantitativně orientovaném pedagogickém výzkumu. *Pedagogická Orientace*, 26(3), 512–536. <https://doi.org/10.5817/pedor2016-3-512>
- Sousa, V. D., & Rojjanasrirat, W. (2010). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17(2), 268–274. <https://doi.org/10.1111/j.1365-2753.2010.01434.x>
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783–805. [https://doi.org/10.1016/s0742-051x\(01\)00036-1](https://doi.org/10.1016/s0742-051x(01)00036-1)
- Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher Efficacy: Its Meaning and measure. *Review of Educational Research*, 68(2), 202–248. <https://doi.org/10.3102/00346543068002202>
- Tsigilis, N., Koustelios, A., & Grammatikopoulos, V. (2009). Psychometric properties of the teachers' sense of efficacy scale within the Greek educational context. *Journal of Psychoeducational Assessment*, 28(2), 153–162. <https://doi.org/10.1177/0734282909342532>
- Valls, M., Bonvin, P., & Benoit, V. (2020). Psychometric properties of the French version of the Teachers' Sense of Efficacy Scale (TSES-12f). *Revue Européenne De Psychologie Appliquée*, 70(3), 100551. <https://doi.org/10.1016/j.erap.2020.100551>
- Yong, E. (2012). Replication studies: Bad copy. *Nature*, 485(7398), 298–300. <https://doi.org/10.1038/485298a>

Martin Fico works at the Faculty of Education of the Masaryk University in Czech Republic. He is involved in research on selected topics in the field of educational policy, such as the prestige of the teaching profession, support for socially disadvantaged pupils, and teaching education.

Contacto: fico@ped.muni.cz