TELL SI

Research Paper

On the Construct Validity of the Persian Version of Ohio State Teacher Efficacy Scale: The Case of Iran

Reza Nejati[1]
Associate Professor, Shahid Rajaee Teacher Training University, Tehran, Iran
Hessam Aldin Sahrapour
M.A. in TEFL, Islamic Azad University of Takestan

Abstract
Teacher efficacy may relate to their effectiveness in terms of class management, instructional strategies, and student achievement. Some questionnaires have been used to capture teachers' efficacy. The most frequently used one is the Ohio State Teacher Efficacy Scale. The study examined the construct validity and reliability of the Persian version of the Ohio State Teacher Efficacy Scale (POSTES) with an available sample of 536 Iranian teachers of English. Their age ranged from 20 to 64 years. The data, ordinal in nature, were submitted to Bayesian confirmatory factor analysis, as computed in AMOS. Convergence Statistic (CS) was 1.0008, which was well below 1.002, the critical value, thereby demonstrating construct validity of the instrument. The reliability of the scale turned out to be high (α = .94). Given the fact that POSTES was both valid and reliable, it is safe to hold that this questionnaire is useful for obtaining information regarding Iranian English teachers' sense of self-efficacy.

---

[1] Corresponding author: reza.nejati@sru.ac.ir

1. Introduction

In conducting any study, the selection of appropriate instruments is one of the important tasks to be done, and choosing improper measures may end in unreliable or even wrong findings, thereby mislead the researchers and end-users of the results. Thus, researchers should be careful in selecting suitable instruments for their investigations.

One area that seems interesting for scholars may be teachers' self-efficacy because teachers are believed to play a key role in the process of education (Hoang, 2018) and their mental health which is critical on their performance and their self-efficacy are two significant factors, among others, which affect them and on how they treat wok stress (Fathi & Derakhshan, 2019). A good number of studies in Iran and other countries has addressed the notion of teachers' self-efficacy (e.g., Bandura, 1977; Dellinger, Bobbett, Olivier, & Ellett, 2008; Doordinejad & Afshar, 2014; Ghonsooly & Ghanizadeh, 2011; Hashemi & Ghanizadeh, 2011; Moradkhani, Rayegan, & Moein, 2017; Morris, Usher, & Chen, 2016; Vaezi & Fallah, 2011; Zonoubi, Eslami Rasekh, & Tavakoli, 2017). Indeed, this concept "has been under study for almost 40 years" (Philippou & Pantziara, 2014, p.96).

Broadly speaking, self-efficacy refers to individuals' beliefs about their capabilities to complete an action successfully (Bandura, 1986) and it is so influential that can predict individuals' motivation (Soodmand Afshar & Hosseini Yar, 2019). In the education setting, apart from some factors such as stress which may influence teachers' health (Harmsen, Helms-Lorenz, Maulana, van Veen, & van Veldhoven, 2019) teachers' self-efficacy appears to be significant for the success of both teachers and students (Klassenet al., 2009; Klassen & Tze, 2014; Zee & Kooman 2016). It may affect teachers' performance, liability, fervor, teaching manners, as well as students' incentive and success (Han & Hiver, 2018; Sun & Wang, 2020; Wolters & Daugherty,

2007; Woodrow, 2011; Zabihi, 2018). Furthermore, teachers' self-efficacy helps to do their best to teach well and helps them generate fresh ideas (Khani & Tarlani Aliabadi, 2016).

Hence, it is crucial to assess teachers' self-efficacy and help those who are less efficacious strengthen their sense of self-efficacy. To meet this end, various instruments have been designed and put to use. Some of them include Character Efficacy Belief Instrument (Milson, 2003), Teacher Self-Efficacy in Behavior Management and Discipline Scale (Emmer & Hickman, 1991), Teacher Self-Efficacy Scale (Friedman & Kass, 2002), The Teacher Efficacy Scale (Gibson & Dembo, 1984), Teacher Self-Efficacy Scale (Schwarzer, Schmitz, & Daytner, 1999), Webb Efficacy Scale (Ashton, Olejnik, Crocker, & McAuliffe, 1982), Bandura's instrument of teacher self-efficacy scale (1997), and Ohio State Teacher Efficacy Scale or OSTES (Tschannen-Moran & Woolfolk Hoy, 2001). It is worth mentioning that another name for OSTES is the Teachers' Sense of Efficacy Scale (TSES).

When it comes to teachers' self-efficacy, the measures mostly chosen by the researchers are questionnaires. In other words, the common point among the articles researching the notion of self-efficacy is the fact that in nearly all of them, a questionnaire has been used to gauge this construct. All of these questionnaires have their pros and cons. This study focused on the widely-used measure in Iran, namely the Persian version of the Ohio State Teacher Efficacy Scale (POSTES).

The reasons to choose this scale to investigate are as follow: First, POTSES embraces prominent dimensions of teachers' self-efficacy. These dimensions are efficacy in student engagement, efficacy in instructional strategies, and efficacy in classroom management; consequently, it can portray a better picture of teachers' self-efficacy. Likewise, Klassen et al., (2009) approved that "a three-factor model of Teacher Sense of Efficacy Scale consisting of self-

efficacy for student engagement, instructional strategies, and classroom management is statistically preferable over a one-factor model of TSE in a range of settings" (p. 75).

The second reason is related to cultural issues. OTSES and its Persian version appear to be appropriate to the Asian context and culture to a large extent. As Klassen et al., (2009) stressed "[our study] provides support for the use of the Teacher Sense of Efficacy Scale outside of culturally Western settings" (p. 75). The need for researching teachers' sense of efficacy beliefs in the EFL context has been recognized because of its unique setting (Chiang, 2008; Faez & Karas, 2017). Faez and Karas (2017) contended that the distinctiveness of the EFL field is echoed in the role of the English language as means of instruction.

Third, the POTSES has been utilized most in studies exploring teachers' self-efficacy. This measure is prevalent and is the most common questionnaire among the studies which have examined teachers' self-efficacy, especially in Iran. POSTES is so popular that in 37 investigations out of 50 studies in 2010-2016 concerning self-efficacy explored by the author, this scale had been employed (e.g., Ghanizadeh & Moafian, 2011; Ghonsooly & Ghanizadeh, 2011; Hashemi Moghadam, 2015; Moradian & Ahmadi, 2014; Sarkhosh & Rezaee, 2014; Veisi, Azizfar, & Gowhary, 2015).

The validity of TSES has been examined in several countries (e.g., Singapore, Korea, Canada, the United States, and Cyprus). In these countries, "the TSES showed convincing evidence of measurement invariance" (Klassen et al., 2009, p. 67). However, its validity and reliability have not been investigated in Iran, and it is not clear whether this scale is valid and reliable in this country or not; because, self-efficacy, like other social factors, is context-specific. As it was mentioned earlier, many researchers in Iran use this measure in their studies. Hence, it is reasonable to see how adequate this

questionnaire is in the Iranian context. To address the adequacy of this scale, its validity and reliability were evaluated.

2. Literature Review

Traditionally, teachers' self-efficacy was estimated by self-report scales. However, most of these scales were general and did not specifically address teachers' qualifications (Philippou & Pantziara, 2014); therefore, more distinct measures were needed (Bandura, 1997, 2006; Goddard, Hoy, & Woolfolk Hoy, 2004; Schunk & Usher, 2011). Of course, it should not be forgotten that the measures should be neither too general nor specific because if they are too general, they will not be able to describe the specific results and if they are too specific, they will violate the generalizability of the outcomes (Schunk & Usher, 2011).

The list of the instruments which have been used to appraise teachers' self-efficacy is long. Mainly, the scales which tried to estimate this concept trace back to two principal theories: social learning theory by Rotter (1966) and social cognitive theory by Bandura (1977) (Tschannen-Moran & Woolfolk Hoy, 2001).

Rotter (1966) was mostly concerned with the extent the teachers could influence and teach the students despite the problems and elements which could affect the learners; the problems which were out of the education setting (Wolters & Daugherty, 2007). Tschannen-Moran and Woolfolk Hoy (2001) explained that the measures which root back to Rotter's theory were composed of Rand measure (Armor et al., 1976), responsibility for student achievement (RSA) (Guskey, 1981), Teacher Locus of Control (TLC) (Rose & Medway, 1981), and the Webb scale (Ashton, Olejnik, Crocker, & McAuliffe, 1982).

The research on teachers' self-efficacy seems to be controversial. Wolters and Daugherty (2007) stated that Rotter (1966) was the first one who tried to pinpoint and measure teachers' self-efficacy. However, some other researchers

On the Construct Validity of …

(Armor et al., 1976; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998) believed that identification of teachers' self-efficacy and investigations on it stem from the works of Rand group in the 1970s with a questionnaire which had only two statements. That questionnaire was chiefly based on the notion that whether teachers have the power to control different environmental parameters or not, and whether the teaching effect lies within their control or is affected by the elements which are external to them (Tschannen-Moran & Woolfolk Hoy, 2001).

Indeed, these two items were assessing two factors: one was the power of teaching compared to external factors such as economic and social issues of the students, their race, and so forth (General Teaching Efficacy) and two, the teachers' beliefs in their capabilities to help the students (Personal Teaching Efficacy). The problem with the Rand measure was its too limited number of segments and hence there were concerns about its reliability. Therefore, researchers sought to design other instruments (Tschannen-Moran & Woolfolk Hoy, 2001).

Hence, responsibility for student achievement was proposed by Guskey (1981). This scale, with 30 items, was by far longer than Rand measure. However, this questionnaire was not easy to use and it was not welcomed by the researchers (Tschannen-Moran & Woolfolk Hoy, 2001).

Simultaneous with Guskey, the scale of teacher locus of control was designed by Rose and Medway (1981). It had 28 items and teachers were required to select between two descriptions of situations to determine responsibility for the students' achievements or setbacks. 14 items were related to the situations concerned with the students' achievements and 14 items were connected to the students' setbacks. In each situation for students' achievements and setbacks, one description attributed it to the teacher and the

other description referred it to the aspects beyond teachers' control (Tschannen-Moran & Woolfolk Hoy, 2001).

According to Rose and Medway (1981), teacher locus of control was a better instrument in measuring teachers' behaviors because of its particularity about teaching situations. Nevertheless, this instrument was not approved extensively by the researchers and was gradually forgotten (Tschannen-Moran & Woolfolk Hoy, 2001).

As mentioned earlier, one of the problems with the Rand measure was its reliability. Therefore, some scholars tried to modify Rand measure items to improve the reliability of the measure, and subsequently, the Webb scale was created. The designers of this scale concluded that the teachers with a higher score on this scale were more patient and calmer in their teaching. However, the Webb scale did not receive much attention and acceptance (Tschannen-Moran & Woolfolk Hoy, 2001).

The second main theory from which the measures of self-efficacy originated was Bandura's social cognitive theory and the instruments were the Ashton vignettes (Ashton, Buhr, & Crocker, 1984), teacher efficacy scale (Gibson & Dembo, 1984), brief eclectic measure (Raudenbush, Rowan, & Cheong, 1992), Bandura's teacher self-efficacy scale (Bandura, 1997), and teacher self-efficacy scale (Schwarzer, Schmitz, & Daytner, 1999) (Tschannen-Moran & Woolfolk Hoy, 2001).

Ashton believed that teacher efficacy is context-specific, to prove her idea, in 1984, Ashton, Buhr, and Crocker prepared 50 items that described situations that a teacher may face. Ashton asked teachers to indicate their abilities in controlling these situations. There were of course two versions of the items: one version checking the effectiveness of teachers alone, and the other one made a comparison between the efficacies of one teacher in comparison with

other teachers. However, this scale was not utilized a lot by the researchers (Tschannen-Moran & Woolfolk Hoy, 2001).

At the same time with Ashton, two researchers named Gibson and Dembo developed a questionnaire based on Rand and Bandura's assumptions. There were 30 items in this questionnaire and it covered Personal Teaching Efficacy (PTE) and General Teaching Efficacy (GTE) (Tschannen-Moran & Woolfolk Hoy, 2001). PTE evaluated self-efficacy and GTE (also called teaching efficacy) weighed outcome expectancy. Outcome expectancy is peoples' belief of the possible results of their actions (Woolfolk Hoy & Burke Spero, 2005).

This measure was prevalent but it suffered from some problems. For example, the issue of confusion: Some items could be considered to both cover PTE and GTE. As a result, although Gibson and Dumbos' (1984) measure was popular but because of some shortcomings such as the accuracy of PTE and GTE, and their fluctuation, the researchers looked for other scales (Tschannen-Moran & Woolfolk Hoy, 2001).

According to Woolfolk Hoy and Burke Spero (2005) a major problem with Gibson and Dembos' measure, was that the second parameter of this scale (teaching efficacy) did not assess the outcome expectancy and hence its name was changed to GTE. In other words, this scale did not go in line with Bandura's (1986) definition of outcome expectancy.

Since scholars were not satisfied with existing measures of teachers' efficacy, they decided to create a scale that was a blending of other measures. For example, Midgley, Feldlaufer, and Eccles (1989) chose five items from Rand measure, two vignettes from academic futility (Brookover et al., 1978), one vignette from the Webb Scale, and developed one item themselves. The resulting scale and the similar ones were combined into one because they were similar; however, it was not quite acceptable because it was claimed that the items of these questionnaires were distinct. Later on, Raudenbush, Rowan, and

Cheong (1992) introduced a measure with just one item, and teachers were required to answer it on a 4-point scale (Tschannen-Moran & Woolfolk Hoy, 2001).

Bandura, who believed teachers' efficacy was not fixed across various tasks and subjects, presented an instrument. In 1997, he designed a measure with 30 items that embraced 7 components of efficacy: decision-making efficacy, school resources efficacy, teaching efficacy, efficacy to engage parental inclusion, efficacy to form a positive school atmosphere, efficacy to engage society inclusion, and disciplinary efficacy. Bandura wished to tailor a scale that involves various aspects of teachers' efficacy. Unfortunately, the present author could not find information regarding the reliability and validity of this scale. This instrument could tackle many of the cases that Bandura aimed to, but the problems were not resolved yet (Tschannen-Moran & Woolfolk Hoy, 2001), and also the available studies that have employed this instrument are scarce (Woolfolk Hoy & Burke Spero, 2005).

In 1999, Schwarzer, Schmitz, and Daytner created a scale with 10 items. They followed Bandura's instructions in designing this measure so that in each statement, they used the word I to evaluate very own teacher's self-efficacy. They also included the verbs *can* or *to be able to* to indicate the mastery expectations. Furthermore, they embedded one barrier in the items as well to distinguish among different levels of teachers' self-efficacy. The problem with this instrument was that it considered teachers' self-efficacy as a one-dimensional concept, whereas self-efficacy has been proven to be a multi-dimensional notion (e.g., Chacon, 2005; Skaalvik & Skaalvik, 2007).

Generally, existing scales do not cover the essential aspects involved in self-efficacy evaluation (Bandura, 1997; Wheatley, 2005). Klassen, Tze, Betts, and Gordon (2011) investigated 218 studies and concluded that nearly half of the measures used in these studies were incompatible and not so related to the

On the Construct Validity of …

concept of self-efficacy beliefs. Likewise, Wyatt (2012) believed that a great number of available instruments had led to imprecise assessments of self-efficacy. These scales assess some *self* but not the self-efficacy beliefs (Bong, 2006).

There are concerns about the interpretation of these scales (Schunk & Pajares, 2009) and the principal question is about the real concept that a measure appraises (Chesnut & Burley, 2015). Moreover, "none of the measures currently in use seems to have found the proper balance between specificity and generality" (Tschannen-Moran et al., 1998, p. 13).

Thereby, the current measures (more specifically, questionnaires) are lagging in estimating teachers' self-efficacy. Over the years, various questionnaires (short and long) have been designed to appraise teachers' self-efficacy. However, according to Tschannen-Moran and Woolfolk Hoy (2001), none of these measures is acceptable to a large extent. Tschannen-Moran and Woolfolk Hoy (2001) held that the attempts that have been made to estimate self-efficacy have not led to a desirable scale and claimed that there are problems with each of the present instruments: problems with their reliability and validity.

Teachers' self-efficacy is a multifaceted construct, and its nature and the issue of forming scales neither too narrow nor too general, have caused developing a proper instrument to be a daunting task. It seems that the current scales have not been able to make a balance between specificity and generality. Furthermore, there are difficulties with the interpretation of the parameter structures and the relationship between the elements found in these measures (Tschannen-Moran & Woolfolk Hoy, 2001).

These scales concentrated more on teachers' ability rather than on their capability; the specificity of these instruments was not high, and they were largely general. In other words, they did not investigate the particular features

of teaching (Henson, 2002). Thus, after so many years of investigating teachers' self-efficacy and designing various instruments, it has been known that there is a need for a novel, valid and reliable scale (Henson, Bennett, Sienty, & Chambers, 2000).

Considering the aforementioned shortcomings of the current scales, Tschannen-Moran and Woolfolk Hoy (2001) decided to tailor a measure to presumably embrace a full range of teachers' self-efficacy facets. Hence, in a seminar at the Ohio State University, different forms for a suitable instrument were examined. Finally, the outline of Bandura's scale was selected. The scholars in that seminar chose the items in Bandura's questionnaire which they thought were salient, and they created some items related to the field of teaching as well which were not present in Bandura's measure (Tschannen-Moran & Woolfolk Hoy, 2001).

In the end, there were more than 100 vignettes. Next, the scholars tested the items one by one and tried to modify them to cover the critical domains of teaching and to reach a final list. Hence, after some discussion and exploration, they came up with 52 items to evaluate all the areas of teaching and teacher abilities (Tschannen-Moran & Woolfolk Hoy, 2001).

The name chosen for this new measure was the Ohio State teacher efficacy scale or Teachers' Sense of Efficacy Scale. The new questionnaire was tested in three studies. After the first study, the number of the items changed to 32, and in the second study, more items were omitted to have 18 statements based on three components of the scale. These three components consisted of efficacy in student engagement, efficacy in instructional strategies, and efficacy in classroom management. In the third research, 18 further items were generated and examined, and at last, the final measure had two forms: a long-form including 24 items, and a short form consisting of 12 items. The validity,

reliability, and factor structure of OSTES were checked as well (Tschannen-Moran & Woolfolk Hoy, 2001).

Tschannen-Moran and Woolfolk Hoy checked the new instrument with many teachers, several teacher educators, and researchers, and according to their feedback, they added more items to the questionnaire and changed some of them. They found out that the classroom management component was weak, and Roberts and Henson (2001) suggested removing this factor, but Tschannen-Moran and Woolfolk Hoy attempted to add more items to the instrument to support this factor instead of deleting it. They also realized that the previous measures had concentrated predominantly on the weak students, thus they made effort to take potential strong students into account too (Tschannen-Moran & Woolfolk Hoy, 2001).

Tschannen-Moran and Woolfolk Hoy (2001) explored the correlation of OSTES with other current scales and noted that "the OSTES could be considered reasonably valid and reliable. With either 24 or 12 items, it is of reasonable length and should prove to be a useful tool for researchers interested in exploring the construct of teacher efficacy" (p. 801). They also emphasized that "the OSTES moves beyond previous measures to capture a wider range of teaching tasks" (p. 801). Besides, they argued that:

> The development of the OSTES is a step forward in capturing what has been an elusive construct. It is superior to previous measures of teacher efficacy in that it has a unified and stable factor structure and assesses a broad range of capabilities that teachers consider important to good teaching, without being so specific as to render it useless for comparisons of teachers across contexts, levels, and subjects (pp. 801-802).

Furthermore, Woolfolk Hoy and Burke Spero (2005) approved the TSES questionnaire and asserted that "the TSES opens new possibilities for research" (p. 354). Marsh and Hau (2004) advocated international analogies and maintained that these comparisons give scholars "a valuable heuristic basis to test the external validity and generalizability of their measures, theories, and models" (p. 59). Along the same lines, Klassen et al. (2009) did international research on TSES and held that "items on the TSES demonstrate internal consistency in a variety of settings … the TSES showed measurement invariance across groups of teachers in similar cultural groups" (p. 75).

Wolters and Daugherty (2007) investigated and confirmed the reliability and validity of OSTES too and mentioned that "this instrument prompts teachers to reflect on their beliefs overall and not about a particular class of students" (p. 184).

Skaalvik and Skaalvik (2007) agreed that TSES was a good measure; however, they noted that there were two limitations with this scale. First, TSES confined teachers' self-efficacy to only three dimensions and second, the hurdles which according to Bandura (1997) were essential to be included in the statements of the questionnaires, were not present in many of the items.

To sum up, there are a lot of instruments to estimate teachers' self-efficacy but the one which is common in most of the studies is OSTES. However, even though the Persian version of this questionnaire has been commonly employed in the investigations carried out in Iran, its adequacy (validity and reliability) has not been examined yet in this setting. Therefore, the present study attempts to investigate the adequacy of this scale in the Iranian context. To achieve the purpose of this research, the following questions were addressed:

1. Is POSTES a valid scale for tapping teachers' sense of efficacy among Iranian teachers of English?

2.   Is POSTES a reliable scale for tapping teachers' sense of efficacy among Iranian teachers of English?

3. Method
3.1 Participants

536 EFL teachers chosen, based on availability, from 52 English language institutes filled out the POSTES questionnaire. 404 teachers were female and 122 teachers were male. The remaining participants did not specify their gender. English institutes from which EFL teachers were chosen were located in Tehran, Garmsar, and Karaj. Their age ranged from 20 to 64 years.

3.2. Instrument

The Persian Version of Ohio State Teacher Efficacy Scale (POSTES), consisting of 24 items, was used. It is of note that eight items were dedicated to each subscale of self-efficacy in this scale. Items 1, 2, 4, 6, 9, 12, 14, 22 were related to student engagement; items 7, 10, 11, 17, 18, 20, 23, 24 connected to instructional strategies; and items 3, 5, 8, 13, 15, 16, 19, 21 pertained to classroom management (Tschannen-Moran & Woolfolk Hoy, 2001). The items were on a nine-point Likert scale ranging from 1 <u>nothing</u> to 9 <u>a great deal</u>.

3.3 Procedure

The present researchers thought that the language barrier could yield construct irrelevant information. Thus, they had to make sure that the participants fully understood the items in the questionnaire. Consequently, they used the Persian version of OSTES to do the study. The translated version was checked with five Persian-speaking university professors of English. To administer the questionnaire, the researcher went to several institutes and handed the questionnaires to the participants.

4. Results and Discussion

As stated before, the purpose of the research was to estimate psychometric adequacy, namely construct validity and reliability of the POSTES in Iran

because self-efficacy like other social issues is context-specific. The POSTES is a 24-item scale that purports to measure three components, namely 'efficacy in student engagement, efficacy in instructional strategies, and efficacy in classroom management. The researchers assumed that POSTES score data were most adequately represented by a covariance structure to examine the association among the variables (i.e., items and components of the POSTES questionnaire).

Given that the items comprising this instrument were based on a nine-point scale, the analyses had to be based on a methodology that takes this ordinality into account. Hence, the data of the present study was submitted to Bayesian confirmatory analysis, as computed in AMOS. AMOS required a path model for confirmatory factor analysis. The path model is displayed in Figure 1 and it was drawn based on the original make-up of the questionnaire:

- Efficacy in student engagement: items 1, 2, 4, 6, 9, 12, 14, 22
- Efficacy in instructional strategies: items 7, 10, 11, 17, 18, 20, 23, 24
- Efficacy in classroom management: items 3, 5, 8, 13, 15, 16, 19, 21

AMOS makes several diagnostic indices available for the users to check the convergence of the Markov Chain Monte Carlo (MCMC) sampling method, namely Convergence Statistics, posterior distribution, regression weights, and covariance indices. These indices are presented below.

Some respondents did not provide information for a limited number of items in the questionnaire. All in all, 69 cells out of 12864 cells were missing, that is % 0.005 of cases. Although Bayesian confirmatory analysis can be run on datasets with missing cases, the data were imputed through the stochastic regression imputation option. For the aim of this study, non-informative priors (Uniform Prior Distribution) were used.
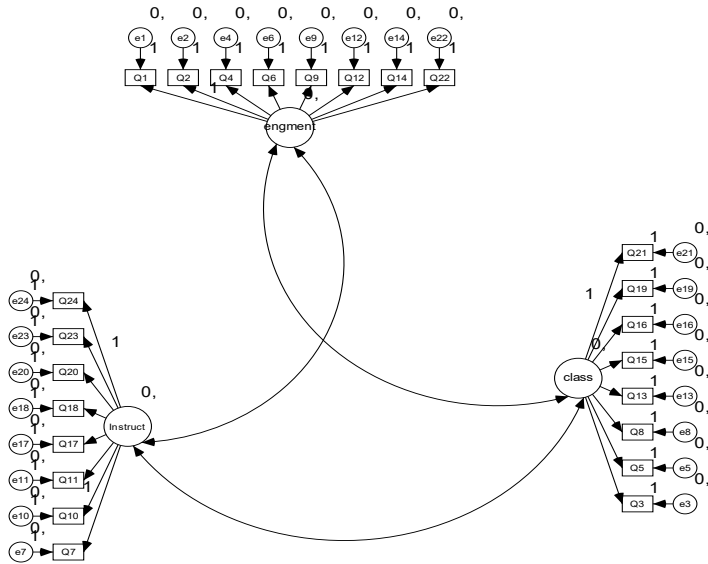
On the Construct Validity of …

Figure 1. Covariance Structure of Persian Version of Ohio State Teacher Efficacy Scale (POSTES)

After (500+68, 501) * 4 times of simulation, the model converged. The largest Convergence Statistic (CS) in this research was 1.0008, which was well below the 1.002 criterion that indicates acceptable convergence by the AMOS default value.

The hypothesized model was tested through analysis of covariance. As stated earlier, the covariance structure (Figure1) examined the relationship between three variables comprising the construct of teacher efficacy, namely Efficacy in student engagement, Efficacy in instructional strategies, and Efficacy in classroom management. Effectively, the model tested the null hypothesis that the covariance between three variables was 0 (Table1)

Table 1
Covariance Structure of Efficacy Variables

Covariance	Mean	S.E.	S.D.	C.S.	95% Lower bound	95% Upper bound
engagement<->Instruct	0.65	0.00	0.07	1.00	0.52	0.80
class<->engagement	0.59	0.00	0.07	1.00	0.46	0.74
class<->Instruct	0.66	0.00	0.06	1.00	0.55	0.78

As it is shown in Table 1, the covariance index between efficacy in student engagement and efficacy in instructional strategies is 0.65, between efficacy in classroom management and efficacy in student engagement is 0.59, and between efficacy in classroom management and efficacy in instructional strategies is 0.66. All mean regression weights exceed 0.5. Hence, it is safe to reject the null hypothesis and conclude that a strong relationship exists between the variables of interest.

The Standard Error (SE) is zero in all covariance indexes. That is to say that the mean covariance among the variables is very precise. The likely distance between the posterior mean and the unknown true parameter is being reported in the third column, labeled SD. The indexes of 0.07 and 0.06 are negligible. Effectively, the Standard Error in maximum likelihood estimation is very marginal. The CS (column 4) compares the variability within parts of the analysis sample to the variability across these parts. A value of 1.00 represents perfect convergence among the variables studied here. In all three variables, the mean covariance lies between 95% Lower bound and 95% Upper bound which is another indication of coherence in the model. All the information presented in Table 1 helps us hold that the current model of teacher efficacy is adequately valid and reliable in the context studied here, namely Iran.

On the Construct Validity of …

Further analysis was done on all items of the questionnaire. The results as displayed in Table 2 suggest that the mean regression of all items weights exceed 0.5. Therefore, it can be claimed that a strong relationship exists among items of the questionnaire and the variables of interest. SE is zero in all cases indicating that the measurement suffers no errors. The likely distance between the posterior mean and the unknown true parameter ranges between 0.06 and 0.11 which is negligible. The CS value of 1.00 represents perfect convergence among the items of the questionnaire examined in this study.

In all items, the mean covariance lies between 95% Lower bound and 95% Upper bound which is another indication of a logical model. All the information demonstrated in Table 2 helps the present researchers to claim that the items comprising the current model of teacher efficacy are adequately valid and reliable in the context of Iran.

In addition to the CS value, one needs to examine the posterior density graphs. Such graphs are available for all items. However, one graph is displayed here.

As Figure 2 shows, the distribution falls to the right of 0 and is almost well-shaped, thereby indicating acceptable convergence. The shape of the distribution for all items of the scale is like this one. Due to space limitations, other plots are not presented here. The complete set of the plots is available upon Email request.

Table 2
Regression Weights for POSTES

Regression weights	Mean	S.E.	S.D.	C.S.	95% Lower bound	95% Upper bound
Q2<--engage	0.97	0.00	0.11	1.00	0.78	1.20
Q4<-- engage	1.16	0.00	0.11	1.00	0.96	1.40
Q6<-- engage	1.10	0.00	0.11	1.00	0.91	1.33
Q9<-- engage	1.17	0.00	0.11	1.00	0.97	1.42
Q12<-- engage	1.12	0.00	0.11	1.00	0.93	1.36
Q14<-- engage	1.19	0.00	0.11	1.00	0.99	1.42
Q22<-- engage	1.11	0.00	0.12	1.00	0.89	1.38
Q10<-- Instruct	0.99	0.00	0.06	1.00	0.87	1.13
Q11<-- Instruct	1.10	0.00	0.07	1.00	0.97	1.24
Q17<-- Instruct	1.06	0.00	0.06	1.00	0.95	1.19
Q18<-- Instruct	1.19	0.00	0.07	1.00	1.06	1.32
Q20<-- Instruct	1.02	0.00	0.06	1.00	0.91	1.13
Q23<-- Instruct	1.21	0.00	0.07	1.00	1.08	1.35
Q19<--class	1.15	0.00	0.06	1.00	1.03	1.28
Q16<--class	0.96	0.00	0.06	1.00	0.85	1.09
Q15<--class	1.19	0.00	0.07	1.00	1.06	1.33
Q13<--class	1.01	0.00	0.06	1.00	0.89	1.13
Q8<--class	0.82	0.00	0.07	1.00	0.70	0.95
Q5<--class	0.78	0.00	0.06	1.00	0.66	0.91
Q3<--class	1.06	0.00	0.07	1.00	0.94	1.20

Group number 1
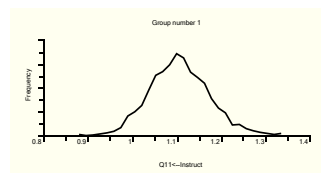Q11<--Instruct

On the Construct Validity of …

Figure 2. Posterior distribution of item 11 to the efficacy of instructional strategies

AMOS also produces additional plots that help us determine the likelihood that the MCMC samples have converged to the posterior distribution via a simultaneous distribution based on the first and last thirds of the accumulated samples.

Looking at Figure 3, readers easily observe that the two distributions are almost identical; consequently, signifying that AMOS has successfully identified important features of the posterior distribution of Item 11. Notice that this posterior distribution appears to be centered at some value near 1.1. The shape of the distribution for all items of the scale is almost like this one. However, due to space limitations, other plots are not presented here. The complete set of the plots is available upon Email request.

To sum up, indices of covariance among variables of this study illustrate that a null hypothesis of covariance structure is not supported here. Therefore, these findings speak well for the validity of our hypothesized structure of the POSTES in Iran.

The reliability of the questionnaire was evaluated in terms of internal consistency, Cronbach's Alpha. The index turned out to be 0.94 and this estimate helps us claim that the scale is highly reliable.
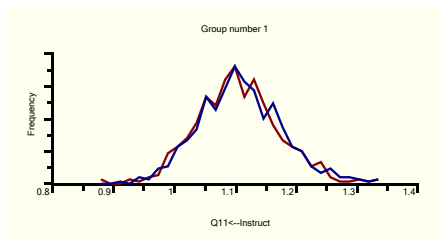

Figure 3. Posterior and simultaneous distribution based on the first and last thirds of the accumulated samples of item 11.

5. Conclusion

In response to researchers who have called for verifying the OSTES in new contexts (Woolfolk Hoy & Burke Spero, 2005 Chiang, 2008; Faez & Karas, 2017 Zonoubi, Eslami Rasekh, & Tavakoli, 2017), the current study examined the validity and reliability of this questionnaire in a new setting, Iran. This investigation can be worthwhile for a series of reasons. The most important one is that OSTES is prevalent and is widely used in studies done in Asia and especially in Iran Ghanizadeh & Moafian, 2011; Ghonsooly & Ghanizadeh, 2011; Hashemi Moghadam, 2015; Moradian & Ahmadi, 2014; Sarkhosh & Rezaee, 2014; Veisi, Azizfar, & Gowhary, 2015). It is popular in Iran and many researchers prefer to employ it in their research. The next reason is its consistency with the Asian culture, and the last one is that the developers of this measure tried to include different features of teachers' self-efficacy into this instrument. These dimensions are teachers' efficacy in student engagement, teachers' efficacy in instructional strategies, and teachers' efficacy in classroom management.

In fact, teacher self-efficacy is multifaceted (Morris & Usher & Chen, 2016; Thompson, & Dooley, 2019; Thompson & Woodman, 2019). It includes efficacy in student engagement (get through to the most difficult students, help students think critically, motivate students, get students to believe they can do well in school work, help students value learning, foster student creativity, improve the understanding of a student who is failing, assist families in helping their children do well in school), efficacy in instructional strategies (respond to difficult questions from students, gauge student comprehension of what has been taught, craft good questions for students, adjust lessons to the proper level for individual students, use a variety of assessment strategies, provide alternative explanations and examples when students are confused, implement alternative strategies in the classroom, provide appropriate challenges for

On the Construct Validity of …

capable students), and efficacy in classroom management (manage most difficult students, control disruptive behavior, establish routines to keep activities running smoothly, make expectations clear about student behavior, establish a classroom management system with each group of students, get children to follow classroom rules, calm a student who is disruptive or noisy, keep a few problem students form ruining an entire lesson, and respond to defiant students).

The results of the analyses confirmed that the Persian version of OSTES could be considered both valid and reliable; thus, the researchers in Iran and other countries with cultures similar to Iran can rest assured regarding the use of this questionnaire. Therefore, it is safe to assert that this scale is a useful measure for researchers interested in obtaining information regarding teachers' sense of self-efficacy.

Readers may agree with the present authors that the OSTES designers have explored the universe of the classroom very well with a sharp lens. All the items relate to the meticulous analysis of the job of teachers all around the world.

It may not be an exaggeration to claim that this questionnaire will generalize to almost all educational institutes in Iran. However, since this study has been carried out with private institutes, the afore-mentioned concepts as conceptualized in POSTES can help researchers in the field of teacher education; adequately appraise English teachers' sense of self-efficacy in Iran and probably in other Persian speaking societies.

This study should point to the need for a psychometrically and theoretically sound measure of the sources of teaching self-efficacy. Such a measure would offer valid information for future studies. As concentration on the sources of teaching self-efficacy grows, so too will the need for a scale that helps the

researchers to evaluate professional development programs (e.g., Henson 2001; Liaw 2009; Ross and Bruce 2007).

The findings of the study hopefully contribute to the self-efficacy beliefs of EFL teachers as a recent branch. According to Hoang (2018), Thompson (2020) and Hoang & Wyat (2020) research into the self-efficacy beliefs of EFL teachers is a newly labeled branch of research into teachers' self-efficacy beliefs. This branch of study has grown out of studies in general education and has increasingly become more discipline-specific.

References

Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). *Analysis of the school preferred reading programs in selected Los Angeles minority schools*. Santa Monica, CA: Rand Corporation.

Ashton, P. T., Olejnik, S., Crocker, L., & McAuliffe, M. (1982). Measurement problems in the study of teachers' sense of efficacy. *Paper presented at the annual meeting of the American Educational Research Association,* New York.

Ashton, P. T., Buhr, D., & Crocker, L. (1984). Teachers' sense of efficacy: A self- or norm-referenced construct? *Florida Journal of Educational Research, 26*(1), 29–41.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bandura, A. (2006). Guide for creating self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich, CT: Information Age Publishing.

Bong, M. (2006). Asking the right question: How confident are you that you could successfully perform these tasks? In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 287–305). Greenwich, CT: Information Age.

Brookover, W., Schweitzer, J., Schneider, C., Beady, C., Flood, P., & Wisenbaker, J. (1978). Elementary school social climate and student achievement. *American Educational Research Journal*, *15*, 301–318.

Chacon, C. T. (2005). Teachers' perceived efficacy among English as a foreign language teachers in middle schools in Venezuela. *Teaching and Teacher Education*, *21*, 257-272.

Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Educational Research Review*, *15*, 1-16.

Chiang, M.-H. (2008). Effects of fieldwork experience on empowering prospective foreign language teachers. *Teaching and Teacher Education, 24*(5), 1270-1287. DOI: 10.1016/j.tate.2007.05.004.

Dellinger, A. B., Bobbett, J. J., Olivier, D. F., & Ellett, C. D. (2008). Measuring teachers' self-efficacy beliefs: Development and use of the TEBS-self. *Teaching and Teacher Education, 24*(3), 751-766.

Doordinejad, F. G., & Afshar, H. (2014). On the relationship between self-efficacy and English achievement among Iranian third-grade high school students. *International Journal of Language Learning and Applied Linguistics World (IJLLALW), 6*(4), 461-470.

Emmer, E., & Hickman, J. (1991). Teacher efficacy in classroom management and discipline. *Educational and Psychological Measurement*, *51*, 755–765.

Faez, F., & Karas, M. (2017). Connecting language proficiency to (self-reported) teaching ability: a review and analysis of research. *RELC Journal, 48*(1), 135-151.

Fathi, J., & Derakhshan, A. (2019). Teacher self-efficacy and emotional regulation as predictors of teaching stress: An investigation of Iranian English language teachers. *Teaching English Language* (*TEL*), *13*(2), 117-143.

Friedman, I. A., & Kass, E. (2002). Teachers self–efficacy: A classroom organization conceptualization. *Teaching and Teacher Education, 18*, 675–686.

Ghanizadeh, A., & Moafian, F. (2011). The relationship between Iranian EFL teachers' sense of self-efficacy and their pedagogical success in language institutes. *Asian EFL Journal, 13*(2), 249-272.

Ghonsooly, B., & Ghanizadeh, A. (2011). Self-efficacy and self-regulation and their relationship: A study of Iranian teachers. *The Language Learning Journal, 41*(1), 1-17.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, *76*, 569–582.

Goddard, R. G., Hoy, W. K., & Woolfolk Hoy, A. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33*(3), 3–13.

Guskey, T. R. (1981). Measurement of responsibility teachers assume for academic successes and failures in the classroom. *Journal of Teacher Education, 32*, 44–51.

Han, J., & Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: The dynamics of change. *Journal of Second Language Writing, 40*, 44–59.

Harmsen, R., Helms-Lorenz, M., Maulana, R., van Veen, K., & van Veldhoven, M. (2019). Measuring general and specific stress causes and stress responses among beginning secondary school teachers in the Netherlands. *International Journal of Research & Method in Education, 42*(1), 91-108.

Hashemi, M. R., & Ghanizadeh, A. (2011). Emotional intelligence and self-efficacy: A case of Iranian EFL university students. *International Journal of Linguistics, 3*(1), 1-16.

Hashemi Moghadam, H. R. (2015). An investigation into the relationship between Iranian high school EFL teachers' emotional intelligence and their self-efficacy. *International Journal of Social Sciences and Education, 5*(3), 509-518.

Henson, R. K., Bennett, D. T., Sienty, S. F., & Chambers, S. M. (2000). The relationship between means-end task analysis and context-specific and global efficacy in emergency certification teachers: Exploring a new model of efficacy. *Paper presented at the annual meeting of the American Educational Research Association*, New Orleans, LA.

Henson, R. K. (2001). The effects of participation in teacher research on teacher efficacy. *Teaching and Teacher Education, 17*, 819–836.

Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist, 37*, 137–150.

Hoang, T. (2018**).** Teacher Self-efficacy Research in English as Foreign Language Contexts: A Systematic Review. *The Journal of Asia TEFL 15*, 976-900.

Hoang & Wyat (2020). Exploring the self-efficacy beliefs of Vietnamese pre-service teachers of English as a foreign language. System, 2021.

Khani, R., & Tarlani Aliabadi, H. (2016). Can learning styles, teaching styles, and teachers' self-efficacy predict Iranian EFL students' language learning achievements? A structural equation modeling analysis. *Teaching English Language* (*TEL*), *10*(2), 95-112.

Klassen, R. B., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wong., I. Y. F., & Georgiou, T. (2009). Exploring the validity of teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology, 34*, 67-76.

Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: a meta-analysis. *Educational Research Review, 12*, 59–76.

Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998– 2009: Signs of progress or unfulfilled promise? *Review of Educational Research, 23*, 21–43.

Liaw, E. C. (2009). Teacher efficacy of pre-service teachers in Taiwan: the influence of classroom teaching and group discussion. *Teaching and Teacher Education, 25*, 176–180

Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*, 56–67.

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self-and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81*(2), 247-258.

Milson, A. (2003). Teachers' sense of efficacy for the formation of students' character. *Journal of Research in Character Education*, *1*(2), 89–106.

Moradian, M. R., & Ahmadi, N. (2014). The relationship between reflectivity in teaching and the sense of self-efficacy among novice and experienced Iranian EFL teachers. *Paper presented at the 2nd National Applied Research Conference on English Language Studies,* Tehran, Iran.

Moradkhani, S., Raygan, A., & Moein, M. S. (2017). Iranian EFL teachers' reflective practices and self-efficacy: Exploring possible relationships. *System, 65*(2017),1-14. DOI: 10.1016/j.system.2016.12.011.

Morris, D. B., Usher, E. L., & Chen, J. A. (2016). Reconceptualizing the sources of teaching self-efficacy: A critical review of emerging literature. *Educational Psychology Review, 29*(4), 795-833.

Philippou, G. N., & Pantziara, M. (2014). Developments in mathematics teachers' efficacy beliefs. In B. Pepin & B. Roesken-Winter (Eds.), *From beliefs to dynamic affect systems in mathematics education: Exploring a mosaic of relationships and interactions* Part of the series *Advances in Mathematics Education* (95-117). Switzerland: Springer International Publishing.

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1992). Contextual effects on the self-perceived efficacy of high school teachers. *Sociology of Education, 65*(2), 150-167.

Roberts, J. K., & Henson, R. K. (2001). A confirmatory factor analysis of a new measure of teacher efficacy: Ohio State Teacher Efficacy Scale. *Paper presented at the annual meeting of the American Educational Research Association*, Seattle.

Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *Journal of Educational Research, 74*, 185–190.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, *80*, 1–28.

Ross, J. A., & Bruce, C. (2007). Professional development effects on teacher efficacy: results of randomized field trial. *The Journal of Educational Research, 101*, 50–60.

Sarkhosh, M., & Rezaee A. A. (2014). How does university teachers' emotional intelligence relate to their self-efficacy beliefs? *Porta Linguarum,* (21), 85-100.

Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 35-54). New York, NY: Routledge.

Schunk, D. H., & Usher, E. L. (2011). Assessing self-efficacy for self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 282-297). New York, NY: Routledge.

Schwarzer, R., Schmitz, G. S. & Daytner, G.T. (1999). *Teacher self-efficacy*. [Online publication]. http://strivetogether.org.

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.

Soodmand Afshar, H., & Hosseini Yar, S. (2019). Investigating the relationship between attitude towards professional development, reflective teaching, self-efficacy, and job performance of Iranian English as a foreign language teachers. *Teaching English Language, 13*(2), 147-179.

Sun, T., & Wang, C. (2020). College students' writing self-efficacy and writing self-regulated learning strategies in learning English as a foreign language. *System, 90*, Article 102221.

Thompson, G. (2020). *Exploring language teacher efficacy in Japan*. London: Multilingual Matters.

Thompson, G., & Dooley, K. (2019). Exploring the key domains where teacher efficacy beliefs operate for Japanese high-school English teachers. *Asia Pacific Education Review, 20*(3), 503-518.

Thompson, G., & Woodman, K. (2019). Exploring Japanese high school English teachers' foreign language teacher efficacy beliefs. *Asia-Pacific Journal of Teacher Education, 47*(1), 48-65.

Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*, 202–248.

On the Construct Validity of …

Tschannen-Moran, M., & Woolfolk Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783-805.

Vaezi, S, & Fallah, N. (2011). The relationship between self-efficacy and stress among Iranian EFL teachers. *Journal of Language Teaching and Research, 2*(5), 1168-1174.

Veisi, Sh., Azizfar, A., & Gowhary, H. (2015). The relationship between Iranian EFL teachers' self-efficacy and gender. *International Research Journal of Applied and Basic Sciences, 9*(1), 140-144.

Wheatley, K. F. (2005). The case for reconceptualizing teacher efficacy research. *Teaching and Teacher Education, 21*(7), 747–766.

Wolters, C. A., & Daugherty, S. G. (2007). Goal structures and teachers' sense of efficacy: Their relation and association to teaching experience and academic level. *Journal of Educational Psychology*, *99*, 181–193.

Woodrow, L. (2011). College English writing affect: Self-efficacy and anxiety. *System, 39*(4), 510–522.

Woolfolk Hoy, A., & Burke Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, *21*, 343–356.

Wyatt, M. (2012). Towards a re-conceptualization of teachers' self-efficacy beliefs: tackling enduring problems with the quantitative research and moving on. *International Journal of Research & Method in Education*, *37*, 166–189.

Zabihi, R. (2018). The role of cognitive and affective factors in measures of L2 writing. *Written Communication, 35*(1), 32–57.

Zee, M., & Kooman, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustments, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*.

Zonoubi, R., Eslami Rasekh, A., & Tavakoli, M. (2017). EFL teacher self-efficacy development in professional learning communities. *System, 66*, 1-12.