BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
http://ojs.bilpublishing.com/index.php/jiep

ARTICLE

Using Hierarchical Generalized Linear Modeling to Examine Contextual Differential Item Functioning: A Validity Study of Teachers' Sense of Efficacy Scale

Jing Zhao[1]* Xiaojing Zou[2] Wenpeng Shang[3]

1. Institutional Research, Old Dominion University, Norfolk, United States
2. Chinese and Culture College, Beijing Normal University, Beijing, China
3. College of Foreign Studies, Jinan University, Guangzhou, China

ARTICLE INFO

Article history:
Received: 10 December 2018
Accepted: 31 January 2019
Published: 22 February 2019

Keywords:
HGLM
Teachers' Sense of Efficacy Scale
DIF

ABSTRACT

The purpose of the study was to further investigate the validity of the instrument used for collecting preservice teachers' perceptions of self-efficacy adapting the three-level hierarchical generalized linear modeling (HGLM) model. To serve the purpose, the study used data collected by the research team which elicited preservice teachers' self-efficacy beliefs using *Teachers' Sense of Efficacy Scale (TSES)*. HGLM were used to analyze the data. Results of the HGLM analyses (at level-two) showed that one item in the scale displayed gender DIF. Another item became DIF item when the context variable was added to the level-two model. However, the effect of the context on the DIF item was not big.

1. Introduction

Differential Item Functioning (DIF)

Differential Item Functioning (DIF) analysis has been and will continue to be a popular topic in the field of measurement. DIF refers to differences in the statistical properties of an item between groups of equal ability. The presence of DIF items on a measurement instrument threatens validity of the interpretation of scores (Angoff, 1993).[2] A similar view was shared by Maller (2001)[15] in claiming that DIF items may pose a threat to validity of test scores and may have serious consequences for groups as well as individuals. The probabilities of getting an item correct are determined not only by the trait that the test claims to measure but also by factors specific to groups, such as differential opportunities to learn or differences in socialization.

There are many different ways to look for DIF, which include but are not limited to Mantel-Haenszel, logistic regression, standardization, and item response theory (IRT).

**Corresponding Author:*
Jing Zhao,
Associate Director, Institutional Research, Old Dominion University, Norfolk, VA 23529*, United States*
Email: kellyzjing@gmail.com.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

In recent years, the nested nature of measurement data, such as items within students, students within schools, has also necessitated the need to utilize multi-level hierarchical linear modeling models to detect DIF. The multilevel model has a lot of advantages, as it could provide more accurate estimates of the standard errors of the parameters in the model. As a result, this kind of approach allows the researchers to investigate the impact of different predictors in the higher level units (e.g., schools, curriculum, and classrooms) and on the lower level units (e.g., students) (Kamata, 2001).[11] Because of these advantages of multilevel models, many studies have been done on detecting DIF using multilevel models (e.g., Kamata, 2001; Lin et al., 2016; Williams & Beretvas, 2006).[11][13][23]These studies used two-level models to investigate DIF items.

As an extension to the two-level models, Cheong (2006)[5]used a three-level model to investigate ethnic-racial DIF for 13 dichotomously-scored items which assessed civic-related language skills. Participants in the study included 2,076 U.S. ninth-grade students in 92 schools. In this model, multiple responses to items on rating civic-related language skills (level 1) were nested within students (level 2) and students were nested within schools (level 3). The analysis for the three-level model consisted of four steps: 1) set up the unconditional model, 2) estimate a conditional model with group membership as predictors, 3) investigate and assess which items exhibit DIF and the patterns, directions, and magnitude of the detected DIF, and 4) include the school-level variable in the DIF screening procedure and repeat Steps 2 and 3.

According to Cheong (2006),[5] it was important to include the contextual sources of DIF into an analysis and the use of a multilevel structure has several advantages. For example, the three-level model could provide a framework which integrates measurement properties and structural relationships; therefore, the framework could be used by researchers to evaluate the psychometric properties of the surveys. It also enables the researchers to model the performance of the items and students with student-, classroom-, and school-level variables. As most large-scale assessment programs have the data with nested structure, the approach is very useful in reality.

Because of its advantages, the three-level model was used to investigate whether items in TSES displayed DIF. Cheong's study applied the three-level model to analyze dichotomously-scored items and the present study applied the model to polytomously-scored items.

Teacher Self-Efficacy Beliefs

In his seminal work, Self-efficacy: The exercise of control, Bandura (1997)[3] first termed self-efficacy as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3). Teachers' efficacy beliefs were defined by Tschannen-Moran et al. (1998)[22] as "teacher's belief in his or her own capability to organize and execute courses of action required to successfully accomplish a specific teaching task in a particular context" (p. 233). Tschannen-Moran and Wolfolk-Hoy (2001)[21] reported that teacher self-efficacy beliefs were related to student outcomes, such as achievement motivation, and the students' own sense of efficacy.
Many studies have focused on the gender differences on preservice teachers' self-efficacy beliefs. Liu's study (2008)[14] showed that female preservice teachers at the early childhood, middle childhood, and secondary licensure levels reported higher self-efficacy beliefs than their male counterparts. However, it was found in Gülten's study (2013)[7] that preservice primary mathematics teachers' math literacy self-efficacy beliefs showed no difference in terms of gender.

The majority of studies on preservice teachers' self-efficacy beliefs have simply compared mean scale differences between female and male preservice teachers, without knowing whether items in a teacher self-efficacy measure function equally across gender. Gender is one of the most important grouping variable for DIF detection for measurement in the domains of self-concept (Yin & Fan, 2003),[24] emotion regulation (Anderson et al., 2016),[1] and learning disabilities (Murray et al., 2015).[17] Only when a measure is invariant across groups, the inference of group differences based on the mean difference is valid (Drasgow, 1987).[6] Therefore, it is necessary to investigate whether items in the preservice teachers' self-efficacy survey display differential item functioning (DIF) across gender.

The purpose of the study was to further investigate the validity of instruments used for collecting preservice teachers' perceptions of self-efficacy adapting the three-level HGLM model described in Cheong's study (2006).[5] The focus of the present study was to investigate whether the polytomously-scored items on the preservice teachers' self-efficacy survey function the same across gender and within different school contexts. Hence, the present study also explored contextual DIF. The research questions were:

What are the psychometric properties (e.g., reliability, construct validity) of the TSES?

Do items on the TSES display DIF by gender?

What is the effect of the third level variable, school context (public universities versus private universities) on gender DIF?

DOI: https://doi.org/10.30564/jiep.v2i1.395

2. Methodology

2.1 Instruments

TSES was constructed by Tschannen-Moran and Woolfolk Hoy (2001)[21] and it has been widely used in measuring teachers' efficacy beliefs. The short form of TSES (see Appendix A) was adopted in the study. The short form of the TSES consists of 12 items which are divided into three factors: efficacy for classroom management (CM), efficacy for student engagement (SE), and efficacy for instructional strategies (IS). The 12 items are rated on a 9-point Likert-type scale ranging from 1 = "nothing" to 9 = "a great deal".

Data in the study were collected by The Ohio State University Teacher Quality Program (TQP) research team in 2007 and 2008. The TQP program was a research initiative that included 50 colleges and universities that provided teacher preparation programs in the State of Ohio. The short form of TSES was used by TQP to assess preservice teachers' self-efficacy beliefs. Preservice teachers included in the study were in the final year of their teacher preparation program and about to graduate.

2.2 Participants

Participants in the study were 1,485 preservice teachers in 45 colleges/universities in the State of Ohio who were pursuing licensure to teach high school in the years of 2006-2007 and 2007-2008. The present study focused on the two cohorts and it was a cross-sectional study. The present study focused on secondary preservice teachers as the number of the female and male was more balanced than elementary preservice teachers in the dataset with eight hundred seven secondary preservice teachers being female (54%) and six hundred seventy eight being male (46%). In addition, 887 were attending the program in public universities (60%) and 598 were in private universities (40%).

2.3 Data analysis

Confirmatory Factor Analysis (CFA)

Cronbach's alpha and confirmatory factor analysis were conducted to answer the first research question, what are the psychometric properties of the TSES. Structural equation modeling takes many factors into account such as: interactions of variables, measurement error, and multiple latent independent variables which could be measured by multiple indicators; therefore, it is more powerful than other statistical approaches such as linear multiple regression (Kline, 1998).[12] LISREL 8.7 (Jöreskog & Sörbom, 2004)[10] was used to conduct the CFA analysis.

Hierarchical Generalized Linear Modeling (HGLM)

A three-level HGLM model (at level two) was applied to answer research question 2, whether the items in the instrument display DIF across gender. To answer research question 3, a third level variable, institution (public vs private universities) was added to the two-level model to see whether the results change. The three-level model was estimated adopting the procedures described in Cheong (2006)[5] which consisted of item responses as Level-1 units, secondary preservice teachers as Level-2 units, and institutions (public vs private universities) as Level-3 units.

Step 1: Estimate the Unconditional Model.

At level 1, for response i for teacher j in school k to item p, the model is:

$$\eta_{mijk} = \pi_{0jk} + \sum_{p=2}^{P} \pi_{pjk} X_{pijk} + \sum_{m=2}^{M-1} D_{mij} \delta_m \quad (1)$$

where η_{mijk} is the log-odds of the probability for the i-th response in category m or lower for teacher j in school k. π_{0jk} are adjusted log-odds of a response on a typical item for teacher j in school k, X_{pijk} are predictor variable representing response i for teacher j in school k to item p, and π_{pjk} are the coefficients for predictors X_{pijk} .The first item in each subscale of the two instruments was used as the reference item. δ_m is a threshold that separates categories m-1 and m. Dmij is an indicator for category m.

At level 2,

$$\pi_{0jk} = \beta_{00k} + \mu_{0jk},\ \mu_{0jk} \sim \mathrm{N}(0, \tau) \quad (2)$$

$$\pi_{0jk} = \beta_{p0k} \text{ for } p = 2, \ldots \mathrm{P}$$

Where β_{00k} is the intercept for school k on the performance of self-efficacy beliefs; μ_{0jk}

is assumed to be normally distributed with mean zero and teacher-level variance τ; π_{0jk}

are assumed to be invariant over teachers.

At level 3,

$$\beta_{00k} = \gamma_{000} + v_{00k},\ v_{00k} \sim N(0, \omega_0) \quad (3)$$

$$\beta_{p0k} = \gamma_{p00} \text{ for } p = 2 \ldots \mathrm{P}$$

Where γ_{000} is the mean level of performance on self-efficacy beliefs; the

random effect, β_{p0k} is assumed to be invariant across universities; and v_{00k} is assumed to be normally distributed with a mean of zero and university-level variance ω.

Step 2: Estimate a Conditional Model and Assess Individual Items.

The conditional model at level 1 was the same as the level 1 unconditional model in equation 2. At level 2, the grouping variable is gender (female=0 and male=1). The level-2 model is:

 DOI: https://doi.org/10.30564/jiep.v2i1.395

$$\pi_{0jk} = \beta_{00k} + \beta_{0k} Group_{jk} + \mu_{0jk},\ \mu_{0jk} \sim N(0,\ \tau) \quad (4)$$
$$\pi_{pjk} = \beta_{p0k} + \beta_{pk} Group_{jk} \text{ for } p = 2 \ldots P$$

Where Groupjk is a predictor of membership of group (female=0 and male=1) for teacher j in school k, β_{00k} is the intercept for school k on the performance of self-efficacy beliefs, β_{0k} is the group difference in self-efficacy beliefs between the reference group and group g for school k, β_{p0k} is the item effect for school k, and β_{pk} is the difference in the item effect between the reference group and group k.

In the level-2 model, if an item p does not display DIF, the difference between two groups in their expected log-odds of correct responses to the item should depend solely on the differences in their levels of self-efficacy beliefs and be equal to zero. If the estimate for an item p is not equal to zero, then the item can be judged to display DIF.

At level 3,

$$\beta_{00k} = \gamma_{000} + v_{00k}, v_{00k} \sim N(0, \omega_0) \quad (5)$$
$$\beta_{0k} = \gamma_{00} + v_{0k}, v_{0k} \sim N(0, \omega_g)$$
$$\beta_{p0k} = \gamma_{p00}$$
$$\beta_{pk} = \gamma_{p0}$$

Where γ_{000} the mean level of performance on self-efficacy beliefs and γ_{00} is the difference in self-efficacy beliefs between the reference group and group g.

Step 3: Enter a School-Level Correlate.

In the level-3 model, institution (public universities vs private universities) was added to the level-two model to see whether the school context has any effect on gender DIF. The level-three model is:

$$\beta_{00k} = \gamma_{000} + \gamma_{001} Z_j + v_{00k}, v_{00k} \sim N(0, \omega_0) \quad (6)$$
$$\beta_{0k} = \gamma_{00} + v_{0k}, v_{0k} \sim N(0, \omega_g)$$
$$\beta_{p0k} = \gamma_{p00} + \gamma_{p01} Z_j$$
$$\beta_{pk} = \gamma_{p0} \text{ for } p = 2, \ldots P$$

where β_{00k} is the intercept for school k, β_{0k} is the group difference in self-efficacy beliefs for school k, β_{p0k} is the item effect for school k, β_{pk} is the difference in the item effect between group g and the reference group. γ_{000} is the grand mean level of performance on self-efficacy beliefs and γ_{00} is the difference in self-efficacy beliefs between the reference group and group g. Zj is the school context (public = 0 and private = 1) of teacher j.

For the level-3 model, the logic of detecting DIF for the level-2 model also applies. But for the level-3 model, the group difference in the item difficulties will be adjusted for various levels of the context variables. The multilevel models were run using HLM 6.04 (Raudenbush et al., 2004)[18] with the full penalized quasi likelihood (PQL). Missing data were handled while running the multilevel analyses.

3. Results

3.1 Descriptive Statistics

Descriptive statistics for TSES items were calculated using SPSS 19.0 and are displayed in Table 1. As reflected in the table, the average responses to these items ranged from 6.73 to 7.97 on a 9-point scale. In general, the preservice teachers were pretty sure about their abilities in managing classroom, engaging students, and using instructional strategies. For the preservice teachers, question 5 got the most agreement (M=7.97, SD=1.17) and question 2 got the least agreement (M=6.73, SD=1.53). That is to say, the secondary preservice teachers were most confident about their ability to craft good questions for their students and they were less but still confident about their skills in motivating students who show low interests in school work.(See Table 1.)

3.2 Confirmatory Factor Analysis for TSES

To answer research question 1, Cronbach's alpha and a three-factor CFA model were run for the instrument. Usually normal theory weighted least square chi-square $\chi 2$, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and goodness of fit index (GFI) are used to assess whether a model is a good fit to the data. Since the $\chi 2$ test is sensitive to the sample size, Shumacker and Lomax (2004)[20] suggest using RMSEA, SRMR, and GFI as indices to test the model fit. Browne and Cudeck (1993)[4]- suggest a RMSEA value between .05 and .08 for a good CFA model fit. According to Hu and Bentler (1999),[9] an SRMR value less than .08 and a GFI value greater than .95 mean a good fit.

The path diagram for the three-factor model is presented in Figure 1. As can be seen in Figure 1, for the latent variable, one estimated factor loading for each of the three–factors (i.e., CM, SE, IS) was fixed to 1. All of the 12 observed variables had significant factor loadings (ranging from .65 to .83), ps < .05, on the three factors. Hence, between 42% and 69% of the variance in teacher self-efficacy beliefs can be attributed to the three subscales and the overall rating. Teacher self-efficacy beliefs were appropriately and adequately assessed by the 12 observed variables in the TSES instrument. For the three-factor model, $\chi 2 = 444.27$, df= 51, p < .001, RMSEA = .072, SRMR = .038, and GFI =.95. Although $\chi 2$ is significant, it is due to the large sample size (Schumacker & Lomax, 2004).[20] The acceptable RMSEA, SRMR, and GFI values showed that the three-factor model was a good fit to the data. The CFA results also showed that conducting DIF analysis within each subscale of TSES would ensure that

DOI: https://doi.org/10.30564/jiep.v2i1.395

Table 1. Descriptive Statistics for TSES

		n	M	SD	Min.	Max.
Self-efficacy belief items						
CM	Control behavior (1)	1475	7.55	1.28	2	9
	Get students to follow rules (6)	1469	7.62	1.21	2	9
	Calm students (7)	1475	7.35	1.28	2	9
	Establish system (8)	1475	7.67	1.25	2	9
	Motivate students (2)	1476	6.73	1.53	1	9
SE	Get students to believe they can do well (3)	1473	7.53	1.21	3	9
	Help students value learning (4)	1472	7.15	1.37	2	9
	Assist families in helping students (11)	1469	6.76	1.56	1	9
	Craft questions (5)	1475	7.97	1.17	3	9
IS	Use assessment strategies (9)	1471	7.82	1.28	1	9
	Alternative explanation (10)	1470	7.90	1.16	2	9
	Alternative strategies (12)	1472	7.48	1.29	1	9
Preservice teachers gender						
	Male	1485	46%	50%	0	1
Institution						
	Private	45	40%	49%	0	1

the assumption of unidimensionality is met.

The reliability index for the overall scale, as well as reliability indices for the three subscales, as measured by Cronbach's alpha, was calculated using SPSS 19.0. The results showed that the reliability index for the overall scale is .905. The reliability index for the first factor, CM (items 1, 6, 7, and 8) is .880. The reliability index for the second factor, SE (items 2, 3, 4, and 11) is .817. The reliability index for the third factor, IS (items 5, 9, 10, and 12) is .802.

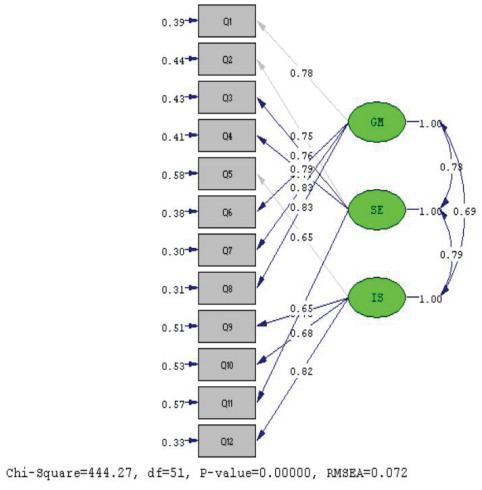

Figure 1. Path Diagram for TSES Three-Factor Model

A three-level HGLM model was used to answer research question 2, whether items in the instrument display gender DIF. To answer research question 3, what is the effect of the school context on gender DIF, institution was entered to the three-level models as a third-level correlate to see whether the results changed.

3.3 Multilevel analyses for TSES subscale 1

The results of the multilevel analyses for TSES subscale 1, CM, are presented in the following tables. For all the four items in this subscale, no teachers chose category 1 as responses. For the analysis, the maximum number of items was 5864, the maximum number of preservice teachers was 1466, and the maximum number of universities was 45.

Step 1: Estimate the Unconditional Model

The results of the unconditional model for subscale 1 are presented in Table 2. The first Thurstone threshold of the reference item, item 1 (How much can you do to control disruptive behavior in the classroom?) was -10.27. In addition, there was significant variance at both the teacher and university levels (6.02 and 0.25 respectively, $p<.001$). Therefore, predictors could be added to the model to account for the variance. Furthermore, distinguishing among teachers across university in self-efficacy beliefs can be

done with a pretty good reliability (0.85).

Table 2. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Unconditional Model for TSES Subscale 1

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.27	0.65	-15.79**
Item 6	γ_{100}	-0.18	0.09	-2.14*
Item 7	γ_{200}	0.48	0.11	4.49**
Item 8	γ_{300}	-0.33	0.06	-5.58**
Threshold 3	δ_3	2.42	0.56	4.36**
Threshold 4	δ_4	3.56	0.60	5.94**
Threshold 5	δ_5	5.65	0.63	9.00**
Threshold 6	δ_6	7.23	0.65	11.17**
Threshold 7	δ_7	9.93	0.65	15.36**
Threshold 8	δ_8	12.04	0.65	18.62**
	Random effects	Variance components		
Teacher-level variance	u_0	6.02**		
University-level variance	v_{00}	0.25**		

Note. df=44 for Intercept. df=5854 for items and thresholds. se=robust standard error. *p<.05. **p<.001.

Step 2: Estimate the Conditional Model and Investigate DIF Items

The results of the conditional model with preservice teachers' gender as the grouping variable are shown in Table 3. The coefficients for all the items in the subscale were not significantly different from zero with p>.05, therefore, all of the teacher self-efficacy belief items did not display significant gender DIF. Furthermore, the teacher-level variance for the model is significant (6.01, p<.001), which means that there is still a lot of variance in preservice teachers' self-belief efficacy beliefs, therefore, other predictors other than gender could be added to the model to account for the variance. The university-level variance is also significant (0.48 with p<.001), indicating that there is still a lot of variance to be explained and university level predictors such as institution could be added to the model to account for the variance.

Table 3. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting DIF for TSES subscale 1

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.26	0.65	-15.82**
TeachGender	γ_{010}	-0.07	0.15	-0.44
Item 6	γ_{100}	-0.27	0.12	-2.18*
Item 6 x TeachGender	γ_{110}	0.18	0.16	1.15
Item 7	γ_{200}	0.57	0.12	4.91**
Item 7 x TeachGender	γ_{210}	-0.19	0.12	-1.55
Item 8	γ_{300}	-0.43	0.09	-5.03**
Item 8 x TeachGender	γ_{310}	0.23	0.14	1.60
Threshold 3	δ_3	2.43	0.55	4.42**
Threshold 4	δ_4	3.57	0.59	6.03**
Threshold 5	δ_5	5.67	0.62	9.14**
Threshold 6	δ_6	7.25	0.64	11.35**
Threshold 7	δ_7	9.96	0.64	15.62**
Threshold 8	δ_8	12.07	0.64	18.95**
	Random effects	Variance Component		
Teacher-level variance	u_0	6.01**		
University-level variance	v_{00}	0.48**		
University-level gender variance	v_{01}	0.13		

Note. TeacherGender=teacher gender (0 = female and 1=male). df = 44 for Intercept and TeachGender. df = 5850 for items and the interaction of items and TeachGender, and thresholds.*p < .05. **p<.001.

Step 3: Enter a University-Level Correlate

The results for the conditional model with institution entered into the model are presented in Table 4. The coefficients for the interactions of institution and item 6 (How much can you do to get children to follow classroom rules? p<.05), of institution and item 7 (How much can you do to calm a student who is disruptive or noisy? p<.05), and institution and item 8 (How well can you establish a classroom management system with each group of students? p<.05) were significantly different from 0, therefore, statistically significant interactions were observed to exist between institution and these items. The positive coefficients for the three items indicated
that preservice teachers at public universities had more

DOI: https://doi.org/10.30564/jiep.v2i1.395

confidence in getting children to follow classroom rules, in calming a student who is disruptive or noisy, and in establishing a classroom management system with each group of students. In addition, item 8 was found to become an item with gender DIF as the coefficient for this item was significantly different from zero ($p<.05$), which indicated that the context had some effects on gender DIF. The large p value (.04) also showed that the magnitude of DIF for this item was not big and the effect of the context on the DIF item was not big. The positive coefficient for item 8 indicated that female preservice teachers felt more confident in establishing a classroom management system than male preservice teachers. Furthermore, no significant interactions were observed by teacher gender and institution interactions. There were statistically significant between-university variations in self-efficacy beliefs, indicating that more predictors other than institution could be added to the model to account for the variance.

Table 4. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting Contextual DIF for TSES subscale 1

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.20	0.63	-16.12**
Insti	γ_{001}	-0.19	0.33	-0.59
TeachGender	γ_{010}	0.10	0.20	0.51
Insti x TeachGender	γ_{011}	-0.45	0.32	-1.41
Item 6	γ_{100}	-0.50	0.15	-3.32*
Item 6 x Insti	γ_{101}	0.54	0.21	2.60*
Item 6 xTeachGender	γ_{110}	0.27	0.21	1.24
Item 6 x InstixTeachGender	γ_{111}	-0.14	0.31	-0.46
Item 7	γ_{200}	0.36	0.15	2.44*
Item 7 x Insti	γ_{201}	0.49	0.23	2.14*
Item 7 xTeachGender	γ_{210}	-0.10	0.13	-0.82
Item 7 x Insti xTeachGender	γ_{211}	-0.17	0.29	-0.57
Item 8	γ_{300}	-0.58	0.08	-7.33**
Item 8 x Insti	γ_{301}	0.35	0.17	2.08*
Item 8 xTeachGender	γ_{310}	0.34	0.17	2.01*
Item 8 x Insti xTeachGender	γ_{311}	-0.25	0.29	-0.86
Threshold 3	δ_3	2 .44	0.54	4.49**
Threshold 4	δ_4	3.58	0.59	6.12**
Threshold 5	δ_5	5.68	0.61	9.27**
Threshold 6	δ_6	7.27	0.63	11.52**
Threshold 7	δ_7	9.98	0.63	15.86**
Threshold 8	δ_8	12.10	0.63	19.25**
	Random effects	Variance component		
Teacher-level variance	u_0	6.01**		
University-level variance	v_{00}	0.48**		
University-level gender variance	v_{01}	0.13		

Note. Insti=Institution (0=public university and 1=private university). df=43 for Intercept, TeachGender, and interaction of TeachGender and Institution. df=5842 for items and the interaction of items and TeachGender and Institution, and thresholds.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

3.4 Multilevel analyses for TSES subscale 2

For the analyses of TSES subscale 2, SE, the maximum number of items was 5844, the maximum number of preservice teachers was 1461, and the maximum number of universities was 45.

Step 1: Estimate the Unconditional Model

The results of the unconditional model for TES subscale 2, SE, are presented in Table 5. The first Thurstone threshold of the reference item, item 2 (How much can you do to motivate students who show low interest in school work?) was -8.53. In addition, there was significant variance at both the teacher (3.69, p<.001) and university level (.24, p<.001). Furthermore, distinguishing among teachers across university in self-efficacy beliefs can be done with high reliability (0.80).

Table 5. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Unconditional Model for TSES subscale 2

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-8.53	0.57	-14.84**
Item 3	γ_{100}	-1.45	0.07	-22.29**
Item 4	γ_{200}	-0.76	0.06	-12.64**
Item 11	γ_{300}	-0.10	0.06	-1.66
Threshold 2	δ_2	1.68	0.61	2.76*
Threshold 3	δ_3	3.54	0.55	6.46**
Threshold 4	δ_4	4.47	0.57	7.90**
Threshold 5	δ_5	6.29	0.59	10.64**
Threshold 6	δ_6	7.70	0.59	13.14**
Threshold 7	δ_7	9.67	0.59	16.27**
Threshold 8	δ_8	11.30	0.60	18.84**
	Random effects	Variance Component		
Teacher-level variance	u_0	3.69**		
University-level variance	v_{00}	0.24**		

Note. df=44 for Intercept. df=5833 for items and thresholds. se=robust standard error. *p<.05. **p<.001.

Step 2: Estimate the Conditional Model and Investigate DIF Items

The results for the conditional model with teachers' gender as the grouping variable are shown in Table 6. The coefficients for all the items in the subscale were not significantly different from zero with p>.05, therefore, all of the teacher self-efficacy belief items did not display significant gender DIF. Furthermore, the teacher-level variance for the model is significant (3.67, p<.001). The university-level variance is also significant (0.35, p<.001).

Table 6. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting DIF for TSES subscale 2

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-8.65	0.57	-15.13**
TeachGender	γ_{010}	0.25	0.14	1.76
Item 3	γ_{100}	-1.50	0.08	-18.16**
Item 3 xTeachGender	γ_{110}	0.10	0.12	0.83
Item 4	γ_{200}	-0.74	0.07	-11.33**
Item 4 xTeachGender	γ_{210}	-0.03	0.10	-0.28
Item 11	γ_{300}	-0.09	0.10	-0.92
Item 11 x TeachGender	γ_{310}	-0.01	0.16	-0.03
Threshold 2	δ_2	1.68	0.61	2.76*
Threshold 3	δ_3	3.54	0.55	6.47**
Threshold 4	δ_4	4.47	0.57	7.91**
Threshold 5	δ_5	6.29	0.59	10.65**
Threshold 6	δ_6	7.70	0.59	13.15**
Threshold 7	δ_7	9.68	0.59	16.29**
Threshold 8	δ_8	11.30	0.60	18.88**
	Random effects	Variance component		
Teacher-level variance	u_0	3.67**		
University-level variance	v_{00}	0.35**		
University-level gender variance	v_{01}	0.06		

Note. TeachGender=teacher gender (0=female and 1=male). df=44 for Intercept and TeachGender. df=5829 for items, the interaction of items and TeachGender, and thresholds. *p<0.05. **p<0.001.

Step 3: Enter a University-Level Correlate

The results of the conditional model with university type entered into the model are presented in Table 7. Significant interactions were observed between university type and item 11 (How much can you assist families in helping their children do well in school? p<.05). Negative coefficient for this item showed that preservice teachers at private universities were more confident about assist-

DOI: https://doi.org/10.30564/jiep.v2i1.395

ing families in helping their children do well in school. In addition, no significant interactions were observed by teacher gender and university type interactions with $p>.05$. Furthermore, there were statistically significant between-university variations in self-efficacy beliefs (3.68, $p<.001$), indicating that more predictors other than university type could be added to the model to account for the variance.

Table 7. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting Contextual DIF for TSES subscale 2

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-8.56	0.61	-13.97**
UniType	γ_{001}	-0.07	0.31	-0.24
TeachGender	γ_{010}	0.36	0.19	1.89
UniType x TeachGender	γ_{011}	-0.29	0.23	-0.91
Item 3	γ_{100}	-1.49	0.07	-21.09**
Item 3 x UniType	γ_{101}	-0.01	0.18	-0.06
Item 3 xTeachGender	γ_{110}	0.07	0.10	0.73
Item 3 x UniTypexTeachGender	γ_{111}	0.07	0.28	0.24
Item 4	γ_{200}	-0.71	0.09	-8.18**
Item 4 x UniType	γ_{201}	-0.08	0.14	-0.63
Item 4 xTeachGender	γ_{210}	-0.11	0.07	-1.68
Item 4 x UniTypexTeachGender	γ_{211}	0.21	0.24	0.87
Item 11 x UniType	γ_{301}	-0.39	0.20	-1.99*
Item 11 xTeachGender	γ_{310}	-0.09	0.18	-0.52
Item 11 x UniTypexTeachGender	γ_{311}	0.18	0.34	0.53
Threshold 2	δ_2	1.68	0.61	2.76*
Threshold 3	δ_3	3.54	0.55	6.47**
Threshold 4	δ_4	4.47	0.57	7.91**
Threshold 5	δ_5	6.30	0.59	10.66**
Threshold 6	δ_6	7.71	0.59	13.17**
Threshold 7	δ_7	9.69	0.59	16.32**
Threshold 8	δ_8	11.31	0.60	18.91**
	Random effects	Variance component		
Teacher-level variance	u_0	3.68**		
University-level variance	v_{00}	0.33**		
University-level gender variance	v_{01}	0.07		

Note. UniType = university type (0 = public university and 1 = private university). df = 43 for Intercept, TeachGender, and interaction of TeachGender and UniType. df = 5821 for items and the interaction of items and TeachGender and UniType, and thresholds. *$p<.05$. **$p<.001$.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

3.5 Multilevel Analyses for TSES Subscale 3

For TSES subscale 3, IS, the maximum number of items was 5864, the maximum number of preservice teachers was 1466, and the maximum number of universities was 45.

Step 1: Estimate the Unconditional Model

The results for the unconditional model for TESE subscale 3, IS, are presented in Table 8. The first Thurstone threshold of the reference item (To what extent can you craft good questions for your students?) was -10.14. In addition, there was significant variance at both the teacher and university levels (3.21 and 0.11 respectively, $p<.001$). Furthermore, distinguishing among teachers across university in self-efficacy beliefs can be done with decent reliability (0.77).

Table 8. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Unconditional Model for TSES subscale 3

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.14	1.22	-8.30**
Item 9	γ_{100}	0.29	0.07	4.15**
Item 10	γ_{200}	0.20	0.07	2.66*
Item 12	γ_{300}	1.05	0.12	9.05**
Threshold 2	δ_2	0.70	0.76	0.93
Threshold 3	δ_3	2.75	1.15	2.40*
Threshold 4	δ_4	3.75	1.20	3.12*
Threshold 5	δ_5	5.76	1.17	4.94**
Threshold 6	δ_6	6.96	1.20	5.81**
Threshold 7	δ_7	8.92	1.19	7.51**
Threshold 8	δ_8	10.60	1.19	8.90**
	Random effects	Variance component		
Teacher-level variance	u_0	3.21**		
University-level variance	v_{00}	0.11**		

Note. df=44 for Intercept. df=5853 for items and thresholds. se=robust standard error. *$p<.05$. **$p<.001$.

Step 2: Estimate the Conditional Model and Investigate DIF Items

The results of the conditional model with preservice teachers' gender as the grouping variable are shown in Table 9. As shown in the table, item 9 displayed significant gender DIF with $p<.001$ and the coefficient for males and females was significantly different from zero. For item 9, the positive coefficient indicated that male preservice teachers found the item harder to endorse and felt less confidence about using different strategies in classroom given the same level of ability as female preservice teachers. Furthermore, the teacher-level variance for the model is significant (3.21, $p<.001$). The university-level variance in the model is also significant (0.13, $p<.05$).

Table 9. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting DIF for TSES subscale 3

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.24	1.23	-8.32**
TeachGender	γ_{010}	0.18	0.10	1.81
Item 9	γ_{100}	0.00	0.10	0.02
Item 9 xTeachGender	γ_{110}	0.62	0.12	5.13**
Item 10	γ_{200}	0.15	0.11	1.35
Item 10 xTeachGender	γ_{210}	0.10	0.16	0.63
Item 12	γ_{300}	1.00	0.16	6.31**
Item 12 x TeachGender	γ_{310}	0.10	0.13	0.75
Threshold 2	δ_2	0.70	0.76	0.93
Threshold 3	δ_3	2.75	1.15	2.40*
Threshold 4	δ_4	3.75	1.20	3.13*
Threshold 5	δ_5	5.77	1.16	4.96**
Threshold 6	δ_6	6.98	1.20	5.83**
Threshold 7	δ_7	8.94	1.18	7.55**
Threshold 8	δ_8	10.62	1.19	8.94**
	Random effects	Variance component		
Teacher-level variance	u_0	3.21**		
University-level variance	v_{00}	0.13*		
University-level gender variance	v_{01}	0.01		

Note. TeachGender = teacher gender (0 = female and 1 = male). df=44 for Intercept and TeachGender. df = 5849 for items , the interaction of items and TeachGender, and thresholds. *$p<.05$. **$p<.001$.

Step 3: Enter a University-Level Correlate

The results of the conditional model with university type entered into the model are presented in Table 10. Item 9 was still found to display gender DIF ($p<.001$) after institution was entered into the model. The context did

DOI: https://doi.org/10.30564/jiep.v2i1.395

not have any effect on the DIF item. No statistically significant interactions were observed between institution and preservice teachers' self-efficacy belief items with the coefficients being not significantly different from zero. In addition, no significant interactions were observed by teacher gender and institution interactions. Furthermore, there were statistically significant between-university variations in self-efficacy beliefs.

Table 10. Fixed Effects (Top) and Variance-Covariance Estimates (Bottom) for the Conditional Model for Detecting Contextual DIF for TSES subscale 3

Parameter	Fixed effects	Coefficient	SE	t Ratio
Intercept	γ_{000}	-10.21	1.29	-7.93**
UniType	γ_{001}	-0.05	0.21	-0.23
TeachGender	γ_{010}	0.20	0.11	1.74
UniType x TeachGender	γ_{011}	-0.04	0.22	-0.20
Item 9	γ_{100}	0.01	0.13	0.11
Item 9 x UniType	γ_{101}	-0.03	0.20	-0.14
Item 9 xTeachGender	γ_{110}	0.57	0.14	4.04**
Item 9 x UniTypexTeachGender	γ_{111}	0.15	0.26	0.56
Item 10	γ_{200}	0.15	0.16	0.91
Item 10 x UniType	γ_{201}	0.01	0.22	0.04
Item 10 xTeachGender	γ_{210}	0.07	0.23	0.32
Item 10 x UniTypexTeachGender	γ_{211}	0.08	0.30	0.26
Item 12	γ_{300}	1.06	0.25	4.27**
Item 12 x UniType	γ_{301}	-0.13	0.30	-0.42
Item 12 xTeachGender	γ_{310}	0.14	0.19	0.72
Item 12 x UniTypexTeachGender	γ_{311}	-0.11	0.26	-0.42
Threshold 2	δ_2	0.70	0.75	0.93
Threshold 3	δ_3	2.75	1.14	2.41*
Threshold 4	δ_4	3.75	1.20	3.13*
Threshold 5	δ_5	5.77	1.16	4.97**
Threshold 6	δ_6	6.98	1.19	5.85**
Threshold 7	δ_7	8.94	1.18	7.57**
Threshold 8	δ_8	10.62	1.18	8.97**
	Random effects	Variance component		
Teacher-level variance	u_0	3.22**		
University-level variance	v_{00}	0.12*		
University-level gender variance	v_{01}	0.01		

Note. UniType = university type (0 = public university and 1 = private university). df = 43 for Intercept, TeachGender, and interaction of TeachGender and UniType. df = 5841 for items and the interaction of items and TeachGender and UniType. *p<.05. **p<.001.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

4. Discussion

Several findings emerged from the present study. As indicated by the descriptive statistics, the secondary preservice teachers in the present study were confident about their abilities in managing classroom, engaging students, and using instructional strategies. The preservice teachers were very confident about their ability inside the classroom, such as crafting good questions for their students and redirecting a student who was disruptive and noisy in the classroom. In addition, they were confident about their ability to motivate students who showed low interests in school work although a student's motivation and performance also depended on his or her home environment. There is some support for the claim that home environment was important for students' motivation. For example, Muola's study (2010)[16] showed that student's motivation to do well in their academic work is to a certain extent dependent on the nature of their home environment.

Results of the CFA analyses provided evidence that TSES had good construct validity. Results showed that the three-factor model for TSES worked well for the present sample in the study. In addition, reliabilities for the overall scales and each subscale are high, which shows that the instrument is a reliable measure of preservice teachers' self-efficacy beliefs.

In terms of DIF items, the results of multilevel HGLM analyses at level-two showed that for TSES subscales 1 (CM) and 2 (SE), no DIF items were detected. For subscale 3 (instructional strategy), item 9 (How much can you use a variety of assessment strategies?) was found to display significant gender DIF. Female secondary preservice teachers reported to be more confident about implementing alternative strategies in the classroom compared to male preservice teachers with the same ability. The reason might be that female preservice teachers practiced using more alternative strategies than male preservice teachers. Further follow-up such as content analysis should be conducted to see whether this item is a biased item. If it is a biased item, then it should be modified or deleted from the instrument.

Item 6 (How much can you do to get children to follow classroom rules?), item 7 (How much can you do to calm a student who is disruptive or noisy?) and item 8 (How well can you establish a classroom management system with each group of students?) in TSES 1were found to have significant interactions with institution. The positive coefficients for the three items showed that secondary preservice teachers who studied to get their licensures at public universities tended to be more confident about "getting children to follow classroom rules", "calming a disruptive or noisy student", and "establishing a classroom management system with each group of students" than their counterparts at private universities. The possible reason is that the teacher education programs at public universities in the State of Ohio were more reputable, therefore, preservice teachers at these universities were more confident about their teaching abilities.

The results of the multilevel analysis also showed that context had effect on DIF items, however, the effect was not big. Some studies have discussed the difference in teacher preparation programs at public versus private universities. For example, Henry et al. (2011)[8] investigated whether teachers got their licensures at public or private universities influenced the achievement of their students. Rosas and West (2011)[19] examined the perceptions of preservice teachers at both public and private universities in the State of Ohio regarding their readiness to teach. However, no research was found on why getting licensures at public or private universities (context) could have influence on preservice teachers' self-efficacy beliefs.

The present study, in the methodological sense, tests the applicability of the three-level model of Cheong (2006)[5] to polytomous data which are more often used in the field of education and psychology. In addition, the present study takes the contextual sources of DIF into account to see whether the context exerts any impact on gender DIF as the context might be one of the sources for DIF items (Cheong, 2006).[5] This multilevel approach has the advantages as evidenced in Cheong's study (2006)[5] as described earlier. The present study also contributes to the self-efficacy literature in that it further explores validity evidence for the widely used self-efficacy survey (TSES) from a new perspective, item bias, which has not been considered before in self-efficacy studies.

One limitation for the study is that due to the unique demographic characteristics of the participants, caution should be exercised when generalizing the results to the preservice teachers at large. The participants in the study were from the colleges and universities which provided licensure programs to preservice teachers in the State of Ohio. All of the participants in the present study were in bachelor's degree program. Around half of them were female and half were male. In addition, half of the preservice teachers were studying at public universities and half at private universities to pursue their licensures. The findings of the study can only be generalized to the population with similar characteristics.

5. Conclusion

To sum up, the present study aimed to investigate validity related issues of the TSES, which measured teachers' self-efficacy beliefs. 1485 preservice teachers who were pursuing their licensures in the colleges and universities

DOI: https://doi.org/10.30564/jiep.v2i1.395

in the State of Ohio in the years of 2006-2007 and 2007-2008 participated in the present study.

The results of the descriptive statistics showed that the preservice teachers were pretty confident about their teaching ability in classrooms, but not so confident about how to motivate students. The results of the CFA analyses showed that the TSES has good construct validity and is a reliable measure for the present sample. The instrument could measure what it is supposed to measure.

In terms of DIF items, one item (item 9) was detected by multilevel model to be DIF items in TSES subscale 3, IS. In addition, item 8 in TSES subscale 1became DIF items when the context variable, institution, was added to the level-two model. Further follow up analysis can be done to investigate whether these items are biased item. Modification could be made to the item or the item could be deleted from the instrument. In general, the four-step procedure (modified to a three-step procedure in the present study) in Cheong's study (2006)[5] worked well for the present sample with high estimate reliability. However, due to some limitations of the present study, caution should be used in generalizing the results.

5.1 Implications for Future Research

The present study is a study using several quantitative techniques, CFA and HGLM analyses to further validate the TSES. The present study has its own significance as it contributes to the sparse literature investigating the validity issues related to measuring preservice teachers' self-efficacy beliefs. The present study also tests the applicability of the three-level model of Cheong (2006)[5] to polymotous data which are more often used in the field of education. However, it also has its limitations as discussed earlier. Therefore, further studies should be done in light of the limitations.

Due to the unique characteristics of the participants in the present study, additional studies using a randomly selected sample from diversified regions are needed to generalize the results. In addition, effect size was not reported in the present study. As effect size is a measure of practical significance and indicates the magnitude of the difference, therefore, in future studies, effect size should be included as part of the results of multi-level model. Also because this is only an empirical study which produced preliminary results with the present sample size, additional simulation studies which simulate different sample size and compare different procedures of DIF estimates should be conducted to have a more holistic result.

In the multilevel analyses, after teacher gender and institution were added to the model, there was still significant variance at the teacher and university level that needed to be explained, which suggested that more predictors could be added to the model. Therefore, further studies could add teacher-level predictors such as teachers' race and university level predictors such as mean SES to the analyses to see whether these variables could help account for the variance in the model. In addition, the results of the studies also showed that the school context had some effect on gender DIF. After the third-level variable (institutions) was added to the second level model, one item became DIF items from DIF-free items. Further studies could be done to show how the context (public universities and private universities) where preservice teachers attend teaching programs influences their self-efficacy beliefs.

Acknowledgement: We thank Dr. Dorinda J. Gallant, The Ohio State University, for her specific instructions and for the time and energy she put in the manuscript.

References

[1] Anderson, L. M., Reilly, E. E , Gorrell, S., Schaumberg, K., & Anderson, D. A. (2016). Gender-based differential item function for the difficulties in emotion regulation scale. Personality and Individual Differences, 92, 87-91.

[2] Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P.W. Holland & H. Wainer (Eds.), Differential item functioning (pp. 3-24). Hillsdale, NJ: Lawrence Erlbaum.

[3] Bandura, A. (1997). Self-efficacy: The exercise of control. New York, NY: W. H. Freeman and Company.

[4] Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen and J. Long (Eds), Testing structural equation modeling (pp.136–62). Newbury Park, CA: Sage.

[5] Cheong, Y. F. (2006). Analysis of school context effects on differential item functioning using hierarchical generalized linear models. International Journal of Testing, 6, 57-79.

[6] Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. Journal of Applied Psychology, 72, 19-29.

[7] Gülten, D. Ç. (2013). An investigation of pre-service primary mathematics teachers' math literacy self-efficacy beliefs in terms of certain variables. International Online Journal of Educational Sciences, 5 (2), 293-408.

[8] Henry, G.T., Thompson, C.L., Bastian, K.C., Fortner, C.K. Kershaw, D.C., Purtell, K.M., & Zulli, R.A. (2011). Does teacher preparation affect student achievement? Manuscripts submitted for Education Finance and Policy.

[9] Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indices in covariance structure analysis:Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1–55.

[10] Jöreskog, K.G., & Sörbom, D. (2004). LISREL 8.7. Scientific Software International, Inc., 1981-2004, IL, U.S.A.

[11] Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. Journal of Educational Measurement, 38, 79-93.

 DOI: https://doi.org/10.30564/jiep.v2i1.395

[12] Kline, R. B. (1998). Principles and practices of structural equation modeling. New York, NY: The Guilford Press.

[13] Lin, P.Y., Childs, R.A., & Lin, YC (2016). Untangling complex effects of disabilities and accommodations within a multilevel IRT framework. Quality & Quantity, 50, 2767-2788.

[14] Liu, J. (2008). A comparison of teacher candidates and first-year teachers by gender and licensure level, in terms of their perceptions of preparation program quality, efficacy beliefs, and concerns about teaching. Unpublished doctoral dissertation, The Ohio State University, USA.

[15] Maller, S. J. (2001). Differential item functioning in the WISC-III: Item parameters for boys and girls in the national standardization sample. Educational and Psychological Measurement, 61, 793-817.

[16] Muola, J.M. (2010). A study of the relationship between academic achievement motivation and home environment among standard eight pupils. Educational Research and Reviews, 5 (5), 213-217.

[17] Murray, A. L., Booth, T., & Mckenzie, K. (2015). An analysis of differential item functioning by gender in the Learning Disability Screening Questionnaire (LDSQ). Research in Developmental Disabilities, 39, 76-82.

[18] Raudenbush, S., Bryk, A., Cheong, Y. F., Congdon, R., du Toit, M. (2004). HLM 6:Hierarchical Linear and Nonlinear Modeling. Lincolnwood: Scientific Software International.

[19] Rosas, C. & West, M. (2011). Pre-service teachers' perception and beliefs of readiness to teach mathematics. Current Issues in Education, 14(1). Retrieved from http://cie.asu.edu/ojs/index.php/cieatasu/article/view/

[20] Schumacker, R. E., & Lomax, R. G. (2004). A beginner's guide to structural equation modeling (2nd ed.). Mahwah, NJ: Erlbaum.

[21] Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: capturing an elusive construct. Teaching and Teacher Education, 17, 783-805.

[22] Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. Review of Educational Research, 68(2), 202-248.

[23] Williams, N.J., & Beretvas, S. N. (2006). DIF identification using HGLM for polytomous items. Applied Psychological Measurement, 30, 22-42.

[24] Yin, P., & Fan, X. (2003). Assessing the factor structure invariance of self-concept measurement across ethnic and gender groups: Findings from a national sample. Educational and Psychological Measurement, 63(2), 296–318.

Appendix A: Teachers' Sense of Efficacy Scale (TSES)

Directions: This questionnaire is designed to help us gain a better understanding of the kinds of things that create difficulties for teachers in their school activities. On a 9 point scale where 1 is "nothing" and 9 is "a great deal," please mark the option corresponding to your opinion about each of the statements below.

	Nothing	2	Very Little	4	Some Influence	6	Quite a Bit	8	A Great Deal
1. How much can you do to control disruptive behavior in the classroom?	m	m	m	m	m	m	m	m	m
2. How much can you do to motivate students who show low interest in school work?	m	m	m	m	m	m	m	m	m
3. How much can you do to get students to believe they can do well in school work?	m	m	m	m	m	m	m	m	m
4. How much can you do to help your students value learning?	m	m	m	m	m	m	m	m	m
5. To what extent can you craft good questions for your students?	m	m	m	m	m	m	m	m	m
6. How much can you do to get children to follow classroom rules?	m	m	m	m	m	m	m	m	m
7. How much can you do to calm a student who is disruptive or noisy?	m	m	m	m	m	m	m	m	m
8. How well can you establish a classroom management system with each group of students?	m	m	m	m	m	m	m	m	m
9. How much can you use a variety of assessment strategies?	m	m	m	m	m	m	m	m	m
10. To what extent can you provide an alternative explanation or example when students are confused?	m	m	m	m	m	m	m	m	m
11. How much can you assist families in helping their children do well in school?	m	m	m	m	m	m	m	m	m
12. How well can you implement alternative strategies in your classroom?	m	m	m	m	m	m	m	m	m

  DOI: https://doi.org/10.30564/jiep.v2i1.395