Assessing Measurement Invariance of Achievement Emotions Questionnaire for Teachers in Prospective Teacher Sample

Sevilay Kilmen (iD)[1,*]

[1]Bolu Abant Izzet Baysal University, Faculty of Education, Department of Educational Sciences, Bolu, Turkey.

Abstract: The purpose of the current study is to determine whether the Achievement Emotions Questionnaire for Teachers (AEQT) is a psychometrically sound instrument to measure prospective teachers' teaching-related emotions. The three-factor model of the AEQT was confirmed in a prospective teacher sample. Also, reliability results showed that the AEQT is a reliable measurement tool. Measurement invariance results revealed that configural, metric, and scalar invariance were provided across gender. These findings support the use of the AEQT when examining differences based on achievement emotions across gender. For teacher training programs, only configural invariance was provided. Although configural invariance suggests that the three-factor structure of the AEQT is the same across the teacher training programs, the lack of metric invariance indicates that the relationship between the items and the underlying latent variable the AEQT factors is not the same across these groups. The observed variables are not related to the latent variable equivalently across teacher training programs. This result does not allow the comparison of path coefficients and covariances between observed and latent variables across teacher training programs. Also, the lack of scalar invariance indicates that different teacher training programs may interpret some items differently and prevent a comparison of averages between these groups.

1. INTRODUCTION

One of the research topics in educational settings is emotions about teaching. Examining the factors influencing emotions about teaching is of considerable importance, given that the investigation of emotions about teaching enables researchers and teacher trainers to better understand and predict prospective teachers' behavior. Indeed, to date, a large body of studies examined the factors related to emotions about teaching in different populations and contexts. The previous research results emphasized that emotions about teaching are related to many important teaching-related factors in educational settings (Henao-Arias et al., 2017), such as burnout (Frenzel et al., 2016), job satisfaction (Moè et al., 2010), teacher-student relationship, classroom discipline, students' engagement (Hagenauer et al., 2015), and self-efficacy (Eren, 2014).

The Achievement Emotions Questionnaire for Teachers (AEQT, Frenzel et al., 2010) is one of the most commonly used instruments for measuring different facets of teachers' achievement

---

*CONTACT: Sevilay Kilmen ✉ kaplansevilay@yahoo.com ⊞ Bolu Abant Izzet Baysal University, Faculty of Education, Department of Educational Sciences, Bolu, Turkey

emotions. The AEQT was used to measure teachers'/prospective teachers' achievement emotions on many different culture samples (Becker et al., 2015; Frenzel et al., 2009; Hong et al., 2016; Klassen et al., 2012). The research results revealed that the three-factor measurement model of the AEQT was confirmed on different culture samples. However, despite the AEQT's widespread use in various countries, studies on testing the measurement invariance of the AEQT across gender and different teacher training programs are missing.

The first aim of the current research is to examine the factor structure of the AEQT by using both parallel analysis and confirmatory factor analysis on a prospective teacher sample. The second aim of the current research is to provide convergent validity evidence by investigating the relationships of prospective teachers' achievement emotions with their professional self-efficacy beliefs on the Turkish prospective teacher sample. The third aim is to examine the measurement invariance of the AEQT across gender and different teacher training programs.

The current study is crucial for three reasons: First, to the author's knowledge, this is the first study to examine if the measurement invariance is established across gender and teacher training programs in emotions about teaching. In previous studies, although participants were compared according to their demographic features, the measurement invariance of the AEQT was not addressed in these studies. If the measurement invariance was not established, it means that comparison groups do not perceive and interpret items in the same way. Therefore, conducting these comparisons may not be proper to see real differences between groups. Examining the measurement invariance of the AEQT provides an evaluation of whether the AEQT measures the same latent construct(s) in different groups (Raykov et al., 2012). Therefore, the results of the current study may be especially useful for studies which compare teachers'/prospective teachers' teaching emotions according to gender and teacher training programs.

Second is that the previous studies investigating emotions about teaching by using the AEQT were often conducted on in-service teacher samples such as German teachers (Becker et al., 2015, Frenzel et al., 2009), Japanese and Korean teachers (Hong et al., 2016), Greek teachers (Karagianni & Papaefthymiou-Lytraand, 2018), and Canadian teachers (Klassen et al., 2012). Studies examining emotions about teaching by pre-service teacher samples were much rarer (e.g., Eren, 2014). The current study provides concrete contributions to the studies which aim to conduct group comparisons by using the AEQT, on a prospective teacher sample, by focusing on evaluating the psychometric quality of the AEQT on a prospective teacher sample.

Third, as mentioned before, previous studies using the AEQT focused on the teacher samples, not prospective teachers. Therefore, convergent validity pieces of evidence were obtained from the teacher sample. Besides measurement invariance, the current study provides supportive evidence for the convergent validity of the AEQT on prospective teacher samples by examining the relationship between prospective teachers' achievement emotion and self-efficacy beliefs.

1.1. Achievement Emotions About Teaching

Achievement emotions were examined in educational settings by dividing them into categories in terms of their features. According to a number of study results, achievement emotions are divided into two primary dimensions as valence and activation. In terms of valance, emotions are divided into two categories: positive versus negative. On the other hand, achievement emotions are classified as activating versus deactivating in terms of activation. For instance, while enjoyment, hope, and pride are positive activating emotions, relief is a positive deactivating emotion. Anger, anxiety, and shame are negative activating emotions while hopelessness is a negative deactivating emotion (Pekrun et al., 2004). In this study, the three-factor structure of AEQT was examined on a prospective teacher sample. According to mentioned explanations about achievement emotions, the AEQT framework focuses on one

positive and two negative activating emotions (i.e., enjoyment, anger, and anxiety) which are prominent achievement emotions for teachers (Frenzel et al., 2009; Sutton, 2004; Sutton & Wheatley, 2003).

Numerous studies have shown that emotions about teaching are the key concepts to closely relate to classroom climate and teaching quality. For example, relevant literature reveals that negative achievement emotions about teaching (i.e., anger and anxiety) are negatively related to key concepts about teaching such as teachers' self-efficacy and enthusiasm (Frenzel et al., 2016, Frenzel et al., 2009; Kunter et al., 2008). Also, teachers' emotions about teaching are related to their perceptions about student characteristics. For example, whereas teachers' perceptions of students' performance, motivation, and discipline during the lessons were positively related to their positive emotions about teaching (i.e., enjoyment), they were negatively related to their negative emotions about teaching (i.e., anger and anxiety, Frenzel et al., 2009). Moreover, teachers' emotions are also closely related to their students' emotions. Indeed, a recent study's results based on longitudinal data demonstrates evidence of the reciprocal transmission of teacher and student emotions (Frenzel et al., 2018).

1.2. Convergent Validity of the AEQT

In this study, to test the convergent validity of the AEQT in the Turkish prospective teacher sample, the relationships between the prospective teachers' emotions about teaching and their professional self-efficacy beliefs were examined. Teacher efficacy belief refers to the "judgment of teachers' capabilities to bring about desired outcomes of students' engagement and learning, even among those students who may be difficult or unmotivated" (Tschannen-Moran & Woolfolk Hoy, 2001, 783). In a number of studies, the relationships between teachers' self-efficacy beliefs and their emotions about teaching were examined (e.g., Hascher & Hagenauer, 2016). Previous study results showed that emotions and self-efficacy are related variables (e.g., Burić et al., 2020). In a study examining the relationships between self-efficacy and emotions about teaching based on a sample who are in the teaching practicum, it was found that prospective teachers' teaching enjoyment in teaching practicum is positively predicted by their self-efficacy, whereas anxiety is negatively predicted (Hascher & Hagenauer, 2016).

On the other hand, teacher self-efficacy was addressed as one latent variable, in some previous studies (e.g., Hascher & Hagenauer, 2016), while in some, this variable was examined in its dimensions (Hagenauer et al., 2015). In the current study, to assess prospective teachers' self-efficacy beliefs, the three-factor teacher self-efficacy beliefs framework described by Tschannen-Moran and Woolfolk Hoy (2001) was adopted. This framework includes three dimensions: self-efficacy for instructional strategies, self-efficacy for classroom management, and self-efficacy for student engagement. Relevant literature shows that specific types of self-efficacy may affect specific teaching emotions (e.g., Hagenauer et al., 2015). Specifically, study results showed that teachers who held high self-efficacy beliefs had more positive emotions (e.g., enjoyment, pride) and less anger and anxiety, compared to teachers who had low self-efficacy beliefs (e.g., Hong et al., 2016). Therefore, by considering that prospective teachers' different types of self-efficacy beliefs may have a different influence on their emotions about teaching, in the current study, the roles of three types of self-efficacy beliefs on emotions about teaching were addressed separately (i.e., instructional strategies, classroom management, and student engagement).

Based on the previous study findings, it is reasonable to hypothesize that the prospective teachers' positive achievement emotion about teaching would positively associate with their self-efficacy beliefs and negative achievement emotion about teaching would negatively associate with their self-efficacy beliefs. Therefore, while testing convergent validity of the AEQT by using prospective teachers' professional self-efficacy, it was expected that the

enjoyment subscale of the AEQT would be related to prospective teachers' self-efficacy beliefs positively and anger and anxiety would be related negatively.

1.3. Measurement Invariance

Measurement invariance is the level of perception and interpretation of scale in the same way across groups (Byrne & Watkins, 2003). When comparing scores obtained from a scale, ensuring measurement invariance between groups is a prerequisite (Marsh et al., 2014). If the scale items are perceived and interpreted differently by the groups, the scores obtained from the comparison of these groups may be misinterpreted. There are four hierarchical types of measurement invariance levels: configural, metric, scalar, and strict invariance (Vandenberg & Lance, 2000).

- Configural invariance: tests if the factor structure of a scale is the same across comparison groups.
- Metric invariance: examines if factor loadings of a scale besides factor structure are equal across comparison groups.
- Scalar invariance: examines if intercepts of a scale besides factor structure and factor loadings are equal across comparison groups.
- Strict invariance: examines if residual variances of a scale besides factor structure, factor loadings, and intercepts are equal across comparison groups.

In the present study, in the configural invariance stage, the AEQT was tested for whether the same factor structure across gender and the teacher training program groups exist. In the metric invariance model stage, the factor loadings of the AEQT items were constrained to be equal across gender and teacher training program groups. In the scalar invariance stage, the AEQT item intercepts were constrained to be equal across the groups in addition to factor loadings. In the last level of measurement invariance procedure, to test strict invariance, the error variances were constrained across groups in addition to the factor loadings and intercepts.

1.4. The Present Study

The purpose of the current study is threefold. The first purpose is to test the factor structure of the AEQT with the Turkish prospective teacher sample. The second purpose is to examine whether the three-factor measurement model of the AEQT had measurement invariance across gender and teacher training programs in the Turkish prospective teacher sample. The third purpose is to provide evidence of the convergent validity of the AEQT in the Turkish prospective teacher sample. The research questions of this study are:

1) Is the factor structure of the AEQT similar to the original scale?
2) Are the configural, metric, scalar, and strict measurement invariance of the AEQT provided across gender?
3) Are the configural, metric, scalar, and strict measurement invariance of the AEQT provided across teacher training program groups?
4) Is there any relationship between prospective teachers' emotions about teaching and their professional self-efficacy, as an indicator of the AEQT?

2. METHOD

2.1. Study Group

To determine the current study's research sample, convenience sampling was used. The general research sample consists of 560 prospective teachers (407 females) majoring in science teaching (n = 107), social sciences teaching (n = 108), English language teaching (n = 138), special education teaching (n = 106), and mathematics teaching (n = 101) in the Faculty of Education of a university located in the north-west of the Black Sea region in Turkey. These were the

participants in the current study. The research sample consists of 133 freshmen, 170 sophomores, 137 juniors, and 120 seniors. Their ages range from 17 to 37 (M = 20.54, S = 2.31).

The research sample was randomly divided into two samples to conduct exploratory and confirmatory factor analyses and examine the convergent validity of the AEQT in two different samples separately. Sample 1 consists of 271 prospective teachers (194 females) who participated in the present research majoring in science teaching (n = 77), social sciences teaching (n = 25), English language teaching (n = 114), and special education teaching (n = 55). There are 73 freshmen, 13 sophomores, 92 juniors, and 93 seniors in Sample 1. Their ages range from 17 to 29 (M = 20.61, S = 1.59).

Sample 2 consists of 289 prospective teachers (213 females) majoring in science teaching (n = 30), social sciences teaching (n = 83), English language teaching (n = 24), special education teaching (n = 51), and mathematics teaching (n = 101). The research sample consisted of 73 freshmen, 13 sophomores, 92 juniors, and 93 seniors. Their ages range from 17 to 29 (M = 20.48, S = 2.81). Measurement invariance analyses were conducted by merging Sample 1 and Sample 2.
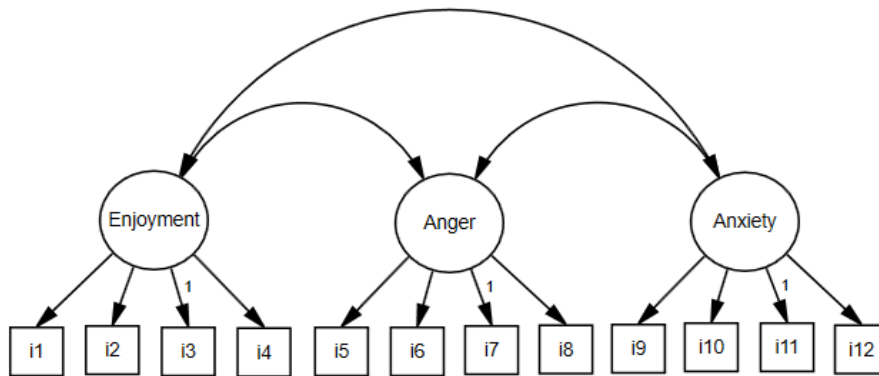
2.2. Research Instruments

The AEQT (Frenzel et al., 2009) and the Ohio State Teacher Efficacy Scale (OSTES, Tschannen-Moran & Woolfolk-Hoy, 2001) were used as measurement tools in the current study. These scales were applied to undergraduate students in the 2018-2019 fall semester. How to answer items in these scales was explained briefly to the participants before the administration, and any questions from the prospective teachers were responded to by the researcher. All participants were informed that their data would not be shared with anyone. All participation was voluntary. The participants completed the scales approximately in 10 min. The general features of the AEQT (Frenzel et al., 2009) and the OSTES (Tschannen-Moran & Woolfolk-Hoy, 2001) were then introduced.

2.2.1. *The achievement emotions questionnaire for teachers (AEQT)*

The AEQT (Frenzel et al., 2010) is a self-report scale with 12 items used to measure teachers' achievement emotions about teaching. The original AEQT was developed to measure in-service teachers' achievement emotions about teaching. The AEQT was adapted to Turkish by Eren (2014). As the AEQT was administered to prospective teachers in Eren's (2014) research, all the AEQT items were converted to the future tense form except for one item (i.e., I feel uneasy when I think about teaching; Eren, 2014). The AEQT consists of three first-order factors (see Figure 1): enjoyment (four items, e.g., I will teach with enthusiasm), anger (four items, e.g., I will get really mad while I teach), and anxiety (four items, e.g., Preparing to teach will cause me to worry). Possible responses range from 1 (strongly disagree) to 5 (strongly agree). The findings showed that all dimensions of the AEQT resulted in satisfactory reliability coefficients (see Table 1).
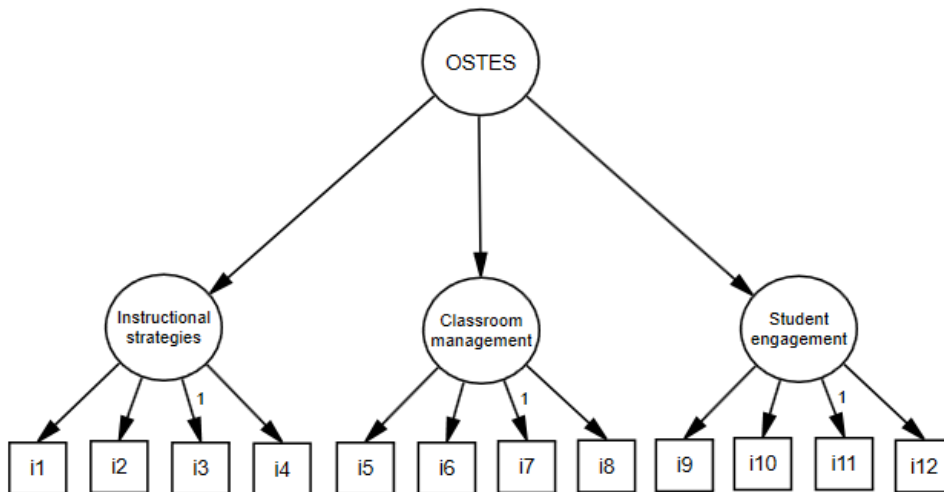
Figure 1. *The AEQT measurement model.*

2.2.2. *The Ohio state teacher efficacy scale (OSTES)*

The OSTES (Tschannen-Moran & Woolfolk-Hoy, 2011) is a self-report scale with 12 items test used to measure teacher self-efficacy. The OSTES consists of one second-order factor and three first-order factors (see, Figure 2): self-efficacy for instructional strategies (four items, e.g., To what extent can you craft good questions for your students?), self-efficacy for classroom management (four items, e.g., How well can you establish a classroom management system with each group of students?), and self-efficacy for student engagement (four items, e.g., How much can you do to get students to believe they can do well in schoolwork?). Prospective teachers responded to items using a 9-point Likert-type scale ranging from 1 (nothing) to 9 (a great deal).

Figure 2. *The OSTES measurement model.*

In the current study, CFA was conducted to see whether the three-factor structure of the OSTES fit the data. CFA results show that the second-order self-efficacy model had a good fit to the current data ($\chi^2_{(51)} = 218.98$; comparative fit index (CFI) = .95; Tucker-Lewis index (TLI) = .94; standardized root-mean-square residual (SRMR) = .06). To evaluate the reliability of the subscales of the OSTES, Cronbach's alpha coefficients were calculated. For all subscales and the whole scale, satisfactory reliability coefficients ranging from .82 to .88 were obtained.

2.3. Data Analysis

Little's missing completely at random (MCAR) test (Little, 1988) was used to examine if missing values are completely at random or not. After non-significant Little's MCAR test

results, the Expectation-Maximization algorithm, which is a technique that uses maximum likelihood estimates for incomplete data, was performed. Original AEQT has a three-factor structure (Frenzel et al., 2010). In order to explore the factor structure of the AEQT on the Turkish prospective teachers, the parallel analysis (Horn, 1965), which is commonly used for scale dimensionality (Timmerman & Lorenzo-Seva, 2011), was conducted on Sample 1. Following the parallel analysis, to examine if the three-factor measurement model of the AEQT was confirmed by the research data, the confirmatory factor analysis was carried out using a different sample (Sample 2) (Kline, 2005). Model fit was evaluated using chi-square (χ^2), comparative fit index (CFI≥.90), and Tucker-Lewis index (TLI≥.90), and standardized root-mean-square residual (SRMR≤.08) (Brown & Cudeck, 1993; Kline, 2005; Hu & Bentler, 1999). To evaluate scales' reliability, Cronbach's alpha was computed.

As the achievement emotions model for teachers obtained from the AEQT was considered to be able to be interpreted differently by the sub-groups in the prospective teacher sample, configural, metric, scalar, and strict measurement invariance across gender and teacher training programs were examined in the present study by using multi-group confirmatory factor analysis. Measurement invariance model comparisons were assessed using ΔCFI cutoff criteria (ΔCFI≤.01; Chen, 2007; Cheung & Rensvold, 2009).
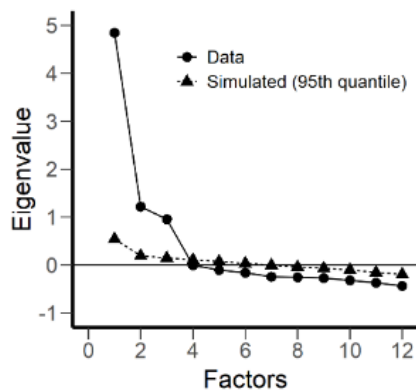
3. RESULTS

3.1. Factor Structure of the AEQT

3.1.1. *Parallel analysis results*

The parallel analysis was applied to the AEQT scores obtained by "Sample 1" (n = 271) to reveal the factor structure of the AEQT in the research sample. The parallel analysis results suggested the three-factor model proposed by Frenzel, Pekrun, and Goetz (2010) ($\chi^2_{(33)}$= 129.123, *p*<.001, see Figure 3 and Table 1). It was found that item loadings on these three dimensions were above .69.

Table 1. *Parallel analysis, confirmatory factor analysis, and reliability results of the AEQT.*

	Parallel analysis results (Sample 1) Factor loadings			Confirmatory factor analysis results (Sample 2) Standardized parameter estimations		Reliability results (Cronbach's alpha) Sample 1	Sample 2	Entire sample
Enjoyment1	**.89**	-.03	.05	.65				
Enjoyment2	**.92**	-.06	-.07	.76		.91	.91	.92
Enjoyment3	**.86**	.00	.00	.69				
Enjoyment 4	**.73**	-.12	.01	.58				
Anxiety1	-.10	**.69**	.06		.78			
Anxiety2	.06	**.91**	-.07		.91	.83	.83	.83
Anxiety3	.04	**.84**	.08		.95			
Anxiety4	-.22	**.69**	.04		.88			
Anger1	.04	.11	**.70**			.44		
Anger2	.02	.04	**.92**			.49		
Anger3	.06	-.03	**.94**			.44	.85	.85	.85
Anger4	-.28	-.03	**.71**			.54		

<div align="center">

$\chi^2_{(33)}$= 129.123, *p*<.001

χ^2_{51}= 259.060, *p*<.001
CFI=.95, TLI=.94,
SRMR=.06

</div>

Figure 3. *Parallel analysis eigenvalues.*

3.1.2. *Confirmatory factor analysis results*

For the confirmatory factor analysis, the three-factor measurement model identified by Frenzel, Pekrun, and Goetz (2010) was used as the baseline for confirmatory factor analysis which was performed on the "Sample 2" (n=289). Each item was specified to reflect the corresponding factor and the three first-order factors were allowed to correlate. When the results were evaluated, it was found that the three-factor measurement model provided a good fit to the data (χ^2_{51}=259.060, CFI = .95, TLI = .94, SRMR = .06, see Table 1), suggesting that the three first-order achievement emotions measurement model offered a reasonably good representation of the data. All standardized factor loadings were above .44. In addition, the findings indicated a moderately latent correlation between anger and anxiety sub-dimensions (r = .495, *p* < .001). These findings suggested that two negative factors of the AEQT can be the first-order factors of a second-order factor. Therefore, a higher-order model was formed. Model 2 comprised one second-order latent factor overarching anger and anxiety, and enjoyment as separate first-order latent factors. But the findings showed that the higher-order model did not yield acceptable fit indexes (χ^2_{32}= 404.324, TLI = .78, CFI = .83, SRMR = .17). Therefore, the higher-order model was rejected. Consequently, the comparison of Model 1 and Model 2 suggested the adoption of Model 1 with three first-order factors (i.e., enjoyment, anger, and anxiety) measured by 12 indicators.

3.2. Measurement Invariance Results

3.2.1. *Measurement invariance across gender*

In the current study, multiple-group confirmatory factor analysis was used to assess configural, metric, scalar, and strict invariance across gender and then across teacher training programs for the three-factor achievement emotion measurement model. The results showed that the configural invariance model across gender presented an acceptable fit to the data (χ^2_{102}= 361.402, CFI = .934, see Table 2). Following the configural invariance model, to examine metric invariance, factor loadings of the AEQT items were constrained to be equal across gender. As seen in Table 2, compared with the configural invariance model as a baseline model, the metric invariance model did not demonstrate any change in CFI (χ^2_{111}= 392.236, CFI = .923, ΔCFI = .000, Chen, 2007; Cheung & Rensvold, 2009). This finding showed that metric invariance was supported in research data. Following the metric invariance model, the scalar invariance model was tested by constraining both item factor loadings and item intercepts to be equal across gender. The results showed that the χ² change between metric and scalar invariance model was not statistically significant, and CFI value did not decrease in the scalar invariance model (χ^2_{120}= 399.709, CFI = .928, ΔCFI = .000). When configural invariance, metric invariance, and scalar invariance models were evaluated together across gender, it was observed that ΔCFI demonstrated no significant reduction in model fit. That is, the AEQT factor

structure, factor loadings, and intercepts did not differ significantly across gender. These findings showed that males and females responded to items of the AEQT in the same way. However, strict measurement invariance implying invariance of residual variances yielded a poor fit to data across gender (χ^2_{132}= 577.709, CFI = .896, ΔCFI = .039). That is, the error variances are different across gender.

Table 2. *Measurement Invariance Results of the AEQT.*

	χ^2	*df*	CFI	ΔCFI
Gender				
Configural	361.402	102	.934	-
Metric	392.236	111	.934	-
Scalar	399.709	120	.935	-
Strict	577.709	132	.896	.039
Teacher training programs				
Configural	616.850	255	.920	-
Metric	720.810	291	.904	.036
Scalar	762.790	327	.903	.001
Strict	968.880	375	.856	.047

3.2.2. *Measurement invariance across teacher training programs*

The same measurement invariance routine was also applied to test measurement invariance for teacher training programs. The AEQT's measurement invariance across science teaching, social sciences teaching, English language teaching, special education teaching, and mathematics teaching programs was investigated. According to the results, the unconstrained configural invariance model data fit was obtained (χ^2_{255}= 616.850, CFI = .920). This finding suggested that the three-factor measurement model is similar across different teacher training programs. Following configural invariance, the metric invariance model was tested by constraining item factor loadings to be equal across teacher training programs. Results showed that the metric invariance model resulted in a significant loss of fit (χ^2_{255}= 720.810, CFI = .904, ΔCFI = .036). The loss of fit results suggested that item factor loadings are different across the teacher training programs. The scalar invariance model implying invariance of intercepts also yielded a poor fit to data, showing that item intercepts are different across teacher training programs (χ^2_{327}= 790.810, CFI = .903, ΔCFI = .001). Lastly, it was found that strict invariance is not supported. That is, the error variances are different across the teacher training programs (χ^2_{375}= 968.880, CFI = .856, ΔCFI = .047).

3.3. Convergent Validity Results

The correlation analysis results showed that the enjoyment component of the AEQT was negatively correlated with anger and anxiety (see Table 3) for both Sample 1 and Sample 2. The convergent validity results showed that, as expected, enjoyment is significantly and positively correlated with instructional strategies, classroom management, and student engagement subscales of the OSTES. Anger and anxiety subscales are significantly and negatively correlated with instructional strategies, classroom management, and student engagement subscales of the OSTES. Of note, the enjoyment subscale of the AEQT, compared with the anger subscale, and anxiety subscales showed a stronger correlation with the OSTES and its subscales.

Table 3. *Convergent validity results of the AEQT.*

	Enjoyment	Anxiety	Anger
Sample 1 (n=271)			
1.OSTES	.40	-.31	-.21
2.Instructional strategies	.33	-.30	-.19
3.Classroom management	.30	-.21	-.07
4.Student engagement	.40	-.29	-.27
Sample 2 (n=289)			
1.OSTES	.49	-.28	-.28
2.Instructional strategies	.42	-.24	-.27
3.Classroom management	.35	-.27	-.11
4.Student engagement	.48	-.22	-.34

4. CONCLUSION

There has been recent interest in examining the impact of teaching-related emotions in teacher training environments. As a result of this interest, researchers need psychometrically-sound items to assess teaching-related achievement emotions. The purpose of the current study was to determine whether the AEQT is a psychometrically sound instrument to measure prospective teachers' teaching-related emotions. From a theoretical point of view, the emerging factor structure implies the existence of highly correlated but also distinct emotions of the AEQT. Indeed, the three-factor model suggested a better fit for the data than did the higher-order model. This finding provides evidence that three teaching-related emotions were distinct. When the reliability results were examined, it was seen that the AEQT is a reliable measurement tool.

The results revealed that configural, metric, and scalar invariance were established across gender. These findings support the use of the AEQT when examining differences based on achievement emotions across gender (Brown, 2006). However, the current analyses suggested that while the AEQT demonstrated configural invariance (equal factor structure) across five teacher training programs, metric invariance (equal factor loadings) was not supported.

Although configural invariance suggests that the three-factor structure of the AEQT is the same across the teacher training programs, the lack of metric invariance indicates that the relationship between the items and the underlying latent variable the AEQT factors is not the same across these groups. That is, the observed variables are not related to the latent variable equivalently across teacher training programs. This result does not allow the comparison of path coefficients and covariances between observed and latent variables across teacher training programs (Chen et al., 2005). Also, the lack of scalar invariance indicates that different teacher training programs may interpret some items differently and prevent a comparison of averages between these groups (Van de Schoot et al, 2012; Vandenberg & Lance, 2000).

Strict invariance was established neither across gender nor across teacher training programs. Establishing configural, metric, and scalar invariance across gender could let the researchers compare the latent variables based on gender. But as strict invariance was not established for gender and the teacher training programs, the latent variables are measured with different amounts of error between groups (Van de Schoot et al., 2012). This result could cause a difference in factor score averages across gender and the teacher training programs even when true values of the underlying construct are the same (Brown, 2006). Therefore, it is important to be careful when the AEQT factor scores are compared across these teacher training programs in future investigations. The convergent validity was supported by results that revealed that

self-efficacy and achievement emotions are significantly and selectively related to each other. Enjoyment as a positive emotion is positively associated with self-efficacy, while anger and anxiety as negative emotions are negatively associated with self-efficacy. These results supported many research results that revealed the relationships between teachers'/prospective teachers' emotions and self-efficacy (Brígidoa et al., 2013; Hascher & Hagenauer, 2016; Moè et al., 2010; Stephanou et al., 2013).

4.1. Limitations

As with most educational studies, the current study has some limitations. First, the prospective teacher sample is unbalanced in terms of gender (most of them are female). It could work with a more balanced sample in terms of gender in future studies. Second, the current study focused only on three main emotions about teaching. In future studies, it could be interesting to examine how teachers'/prospective teachers' other emotions about teaching are influenced by gender and self-efficacy beliefs.

Third, in the present study, all analysis which was conducted was limited to 560 pre-service teachers majoring in science teaching, social sciences teaching, English language teaching, special education teaching, and mathematics teaching. The results of the current study showed that metric, scalar, and strict invariance of the AEQT was not provided across teacher training programs. A possible reason of these results may be sample size. Because, in the current study, to examine measurement invariance across teacher training programs, study sample was divided into five categories. Therefore, the number of participants in each category was highly decreased. Model-data fit measures for metric, scalar, and strict invariance of the AEQT may have been decreased depending on this reason. To increase generalizability of research findings and to reassess the measurement invariance, the AEQT can be used in larger and different samples. That is, future studies are needed to cross-validate the results with other samples.

Finally, given that those prospective teachers in the current study have high enjoyment and low anger and anxiety, the current study did not examine the specific relationships between self-efficacy and emotions about teaching in a sample that has low enjoyment and high anger and anxiety. However, it could be helpful to gain a deeper understanding of the relationships between professional self-efficacy and emotions about teaching in future studies using a homogeneous sample.

Declaration of Conflicting Interests and Ethics

The author declares no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author.

ORCID

Sevilay Kilmen ⓘ https://orcid.org/0000-0002-5432-7338

5. REFERENCES

Becker, E. S., Keller, M. M., Goetz, T., Frenzel, A. C., & Taxer, J. (2015). Antecedents of teachers' emotions in the classroom: An intraindividual approach. *Frontiers in Psychology, 6*, 1-12.

Brígidoa, M., Borrachero, A. B., Bermejo, M. L., & Mellado, V. (2013) Prospective primary teachers' self-efficacy and emotions in science teaching. *European Journal of Teacher Education, 36*(2), 200-217.

Brown, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In: K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Beverly Hills, Sage.

Brown, T. (2006). *Confirmatory factor analysis for applied research.* The Guilford Press.

Burić, I., Slišković, A., & Sorić, I. (2020). Teachers' emotions and self-efficacy: A test of reciprocal relations. *Frontiers in Psychology, 11*, 1650.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*(2), 155-175.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modelling, 14*(3), 464-504.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*(3), 471-492.

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

Eren, A. (2014). Relational analysis of prospective teachers' emotions about teaching, emotional styles, and professional plans about teaching. *Australian Educational Researcher, 41*(4), 381-409.

Frenzel, A. C., Goetz, T., Stephens, E. J., & Jacob, B. (2009). Antecedents and effects of teachers' emotional experiences: An integrated perspective and empirical test. In P.A. Schutz & M. Zembylas (Eds.), *Advances in teacher emotion research: The impact on teachers' lives* (pp. 129-152). Springer.

Frenzel, A. C., Pekrun, R., & Goetz, T. (2010). *Achievement emotions questionnaire for teachers (AEQ-T teacher)-user's manual.* Program of Psychology, University of Munich.

Frenzel, A.C., Pekrun, R., Goetz, T., Daniels, L.M., Durksen, T.L., Becker-Kurz, B., & Klassen, R.M. (2016). Measuring teachers' enjoyment, anger, and anxiety: the teacher emotions scales (TES). *Contemporary Educational Psychology, 46,* 148-163.

Frenzel, A. C., Becker-Kurz, B., Pekrun, R., Goetz, T., & Lüdtke, O. (2018). Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment. *Journal of Educational Psychology, 110*(5), 628-639.

Hagenauer, G., Hascher T., & Volet, E. (2015). Teacher emotions in the classroom: associations with students' engagement, classroom discipline, and the interpersonal teacher-student relationship. *European Journal of Psychology of Education, 30*, 385-403.

Hascher, T., & Hagenauer, G. (2016). Openness to theory and its importance for preservice teachers' self-efficacy, emotions, and classroom behavior in the teaching practicum. *International Journal of Educational Research, 77*, 15-25.

Henao-Arias, J. F., Marin-Rodriguez, A. E., & Vanegas-Garcia, J. H. (2017). Education hinging on emotions: An emotional view of education. *Educación y Educadores [online]. 20*(3), 451-465.

Hong, J., Nie, Y., Heddy, B., Monobe, G., Ruan, J., You, S., & Kambara, H. (2016). Revising and validating achievement emotions questionnaire-teachers (AEQ-T-T). *International Journal of Educational Psychology, 5*(1), 80-107.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.

Hu, L. & Bentler, P. (1999). Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Karagianni E., Papaefthymiou-Lytra S. (2018). EFL Teachers' Emotions: The driving force of sustainable professional development. In A. J. Martínez (Ed.), *Emotions in Second Language Teaching* (pp. 385-401). Cham: Springer.

Klassen, R. M., Perry, N. E., & Frenzel, A. C. (2012). Teachers' relatedness with students: An underemphasized component of teachers' basic psychological needs. *Journal of Educational Psychology, 104*(1), 150-165.

Kline, R. B. (2005). *Principles and practice of structural equation modeling.* Guilford Publications, Inc.

Kunter, M., Tsai, Y. M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction, 18*(5), 468-482.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198-1202.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10,* 85-110.

Moè, A., Pazzaglia, F., & Ronconi, L. (2010). When being able is not enough. The combined value of positive affect and self-efficacy for job satisfaction in teaching. *Teaching and Teacher Education, 26*(5), 1145-1153.

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter, S. (2004). Beyond test anxiety: Development and validation of the Test Emotions Questionnaire (TEQ). *Anxiety, Stress & Coping, 17(*3), 287-316.

Pitkäniemi, H. (2017). A teacher's practical theories, self-efficacy, and emotions - What connections do they have, and how can they be developed? *Nordisk Tidskrift för Allmän Didaktik, 3*(1), 2-23.

Raykov, T., Marcoulides, G. A., & Li, C-H. (2012) Measurement invariance for latent constructs in multiple populations: A critical view and refocus. *Educational and Psychological Measurement, 72*(6), 954-974.

Stephanou, G., Gkavras, G., & Doulkeridou, M. (2013). The role of teachers' self- and collective-efficacy beliefs on their job satisfaction and experienced emotions in school. *Psychology, 4*(3A), 268-278.

Sutton, R. E., & Wheatley, K. F. (2003). Teachers' emotions and teaching: A review of the literature and directions for future research. *Educational Psychology Review, 15*(4),327-358.

Sutton, R.E. (2004). Emotional regulation goals and strategies of teachers. *Social Psychology of Education 7*(4), 379-398.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209-220.

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

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486-492.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.