Education and Science tedmem

Vol 39 (2014) No 176 31-50

Development of Classroom Assessment Environment Scale (CAES): Validity and Reliability Study

Mustafa İlhan [1], Bayram Çetin [2]

Abstract

In this research, we aimed to develop an instrument that could be used to measure students' perceptions of the classroom assessment environment in a valid and reliable manner. The research was carried out in spring 2013–2014 with a total of 800 students who constituted four separate study groups. In this study, expert opinions were obtained to determine the scale's content validity and face validity, while exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were performed to assess the construct validity of the scale's measures. The EFA yielded a construct that consisted of 18 items and two factors that explained 31.52% of the total variance. These factors were named as follows: Learning-Oriented Assessment Environment (LOAE) and Performance-Oriented Assessment Environment (POAE). Findings obtained from the CFA demonstrated that the construct with 18 items and two factors had adequate fit indexes. The reliability of the measures obtained using the LOAE and POAE subscales was examined via Cronbach's alpha, composite reliability, and the test–retest method, which produced reliability coefficients that fell within acceptable limits. With the aim of determining the items' discriminatory power, the adjusted item total correlation was examined, and 27% sub-upper group comparisons were made. The findings obtained from the item analysis showed that all of the items on the scale were discriminatory. Based on these findings, it could be concluded that the scale is an instrument that produces valid and reliable measures, and that can be used to determine students' perceptions of the classroom assessment environment.

Keywords

Classroom assessment environment
Scale development
Validity
Reliability

Article Info

Received: 05.06.2014
Accepted: 12.01.2014
Online Published: 12.16.2014

DOI: 10.15390/EB.2014.3334

Introduction

The social context is regarded as one of the essential factors that shape a person's behaviors. In early research studies that have investigated the influence of social context on an individual's behaviors, the importance of family context was highlighted. In addition to this, there has been evidence pointing to the critical role played by the classroom environment on an individual's

[1] Dicle University, Ziya Gökalp Faculty of Education, Primary Education Department, Turkey, mustafailhan21@gmail.com
[2] Gaziantep University, Gaziantep Faculty of Education, Department of Educational Sciences, Turkey, bcetin27@gmail.com

behaviors, dating back to the 1930s (Bartha, Dunlap, Dane, Lochmanb & Wells, 2004). Classroom context is defined as a multidimensional environment where all learning activities are realized (Banks, 2012; Kurt, Ekici, Gokmen, Aktas & Aksu, 2013), and it significantly influences students' academic achievement (Wang, Haertel & Walberg, 1994; Wannarka & Ruhl, 2008), their preferences related to learning approaches (deep or surface learning approaches) (Dart et al., 1999; Yuen-Yee & Watkins, 1994), the development of their self-respect (Harbaugh & Cavanagh, 2012), their goal achievement orientations (learning approach, learning avoidance, performance approach, or performance avoidance) that they have adopted (Lau & Lee, 2006; Phan, 2008; Popilskis, 2013), their attitudes, their learning motives, and their rate of learning (Kose & Kucukoglu, 2009). Research on how students' perceptions relate to the classroom context and influence their cognitive and affective characteristics has revealed that a great majority of variance related to learning output can be explained by perceptions related to the classroom environment (Dorman, 2001). The classroom environment is highly comprehensive in that it covers all stages of learning – from planning to the completion of the learning – teaching process. The classroom environment incorporates the rules that are applied in the classroom, the clarity of these rules, how they were formed, the communicative environment in the classroom, students' expectations, the methods and techniques used in the learning process (Kurt et al., 2013), innovation and variety in these methods and techniques, students' participation in the decision-making processes in the classroom (Mesa, 2012), the classroom's physical environment, teachers' and students' characteristics, as well as the classroom climate (the social, psychological, and emotional interactions that occur in class). One of the significant components of the classroom environment is the class assessment environment (Banks, 2012). Considering the fact that the majority of classroom time is devoted to assessment-related activities (Mertler, 2003), the classroom assessment environment is considered to be one of the most significant components of the classroom setting (Brookhart & DeVoge, 1999).

Classroom Assessment Environment

The concept of the classroom assessment environment was first proposed by Stiggins and Conklin (1992) based on their observations on teachers' assessment practices. Stiggins and Conklin (1992) defined the classroom assessment environment as a structure consisting of eight components: *i)* assessment objectives, *ii)* assessment methods, *iii)* criteria used in the selection of assessment criteria, *iv)* the quality of assessments, *v)* teachers' background in performing assessments, *vi)* feedback given by teachers, *vii)* teachers' perceptions of students, and *viii)* assessment policies (Griffin, 2009). Of these eight components, seven are directly under the teacher's control (Brookhart & Bronowicz, 2003). Teachers are not entitled to arrange evaluation policies; however, evaluation policies affect teachers' assessment and evaluation practices. Assessment and evaluation choice among teachers mostly depends upon evaluation policies, which have a great impact on curriculum. Therefore, it could be argued that Stiggins and Conklin (1992) prioritized teachers' assessment practices over students' perceptions when conceptualizing the classroom assessment environment (Brookhart, 2003; Brookhart & Durkin, 2003). Following Stiggins and Conklin (1992), Brookhart (1997) redefined the classroom assessment environment by synthesizing the literature from social cognitive theory on motivation and assessment environment (Alkharusi, 2013). Brookhart (1997) defined the classroom assessment environment as the perception of the classroom environment created by students, as determined by assessment objectives, assessment tasks, performance measures used in assessment, and feedback determined by the teacher (Alkharusi, Aldhafri, Alnabhani & Alkalbani, 2014). In other words, students' perceptions of classroom assessment practices constitute the classroom assessment environment (Brookhart & DeVoge, 1999).

The classroom assessment environment has a decisive influence on students' beliefs about their self-efficacy (Alkharusi, 2009), achievement goal orientations (Wang & Cheng, 2010), and motivational beliefs and academic achievements (Alkharusi, 2007). Ames (1992) asserted that students' motivations would be positively influenced by an assessment environment in which *i)* assessment methods that involve diversity, innovation, and active participation are used, *ii)* students' individual progress is prioritized over their performance, as compared to their peers (social comparisons), *iii)* personal assessments are prioritized over general assessments, *iv)* learning-related efforts are more emphasized than the final outcome, *v)* opportunities that enable the student to progress are given, *vi)* students are endowed with an view that sees mistakes as an integral part of the learning process, and *vii)* students are given the opportunity to choose or make decisions regarding the assessment process. McMillan and Workman (1998) also commented on how an assessment environment should be, so that learning outcomes would be positively influenced. According to McMillan and Workman (1998); an assessment environment in which *i)* students are given medium-level-difficulty exams and assessment tasks, *ii)* several assessment methods are used together instead of a single one, *iii)* the assessment criteria are clearly defined and students are informed about these criteria before the assessment practices take place, *iii)* students are given meaningful feedback after their assessments, and *iv)* students' mistakes are seen as instruments that will help them overcome their learning-related problems.

Dimensions of the Classroom Assessment Environment

Students' perceptions of the classroom assessment environment are operationally accounted for within two dimensions: the Learning-Oriented Assessment Environment (LOAE) and the Performance-Oriented Assessment Environment (POAE) (Alkharusi, 2011). The assessment environments in which assessment-related tasks are of a moderate level of difficulty, the assessment criteria are clearly defined, students are given feedback as to how to identify and eliminate deficiencies, mistakes are seen as natural components of the learning process, and students are given the chance to correct their mistakes create learning-oriented assessment perceptions among students. On the other hand, performance-oriented assessment perceptions are created among students by assessment environments where assessment-related tasks are difficult, exam grades are prioritized over learning, performance is prioritized over effort, and social comparisons are taken as measures for success (Alkharusi, 2010a). Students' levels of learning- and performance-oriented assessment perceptions affect their beliefs of their self-efficacy, achievement goal orientations, and academic performances. For instance, while higher performance-oriented perceptions lead to lower self-efficacy beliefs, higher learning-oriented perceptions lead to higher self-efficacy beliefs (Alkharusi, 2009).

Individual and General Perceptions of the Classroom Assessment Environment

Students in a single classroom might develop different and/or similar perceptions of their classroom assessment environment (Alkharusi, 2010b). When the studies conducted to determine students' perceptions of the classroom assessment environment are reviewed, it is evident that some of these studies (Alkharusi, 2009; Church, Elliot, & Gable, 2001) rely on individual perceptions. Researchers who argue that perceptions of the classroom assessment environment should be examined on an individual basis suggest that students in the same classroom might develop diverse perceptions due to the diversity of their individual characteristics and experiences (Ames, 1992). On the other hand, it has been argued in other studies that differences among students' characteristics and experiences do not pose an obstacle for them to develop various shared perceptions of the classroom assessment environment (Church, Elliot, & Gable, 2001). Based on this view, it could be argued that mean scores that reflect the general condition in a classroom can be used, along with individual responses, while examining perceptions of the classroom assessment environment.

In summary, students' perceptions of the classroom assessment environment are closely linked with the learning–teaching processes and outcomes that are produced via these processes. In this respect, it is of great importance to contribute to the relevant literature with a scale that can be used in measuring students' perceptions of the classroom assessment environment. The literature shows that there exists a non-Turkish scale that can be used to measure students' perceptions of the classroom assessment environment (Alkharusi, 2009; 2011). However, no assessment instrument that could serve the same purpose has been found in the relevant literature in Turkish. Therefore, there exists a need to contribute to the Turkish literature by establishing a scale that can be used to measure students' perceptions of the classroom assessment environment. In this study, we aimed to develop the Classroom Assessment Environment Scale (CAES).

Method

Study Group

The study was carried out in spring 2013–2014 with four separate study groups consisting of high school students. The first group comprised 314 students, 153 (48.73%) of whom were female and 161 (51.27%) of whom were male, from three high schools in Batman city center. Of these students, 116 (36.94%) were ninth-graders, 113 (35.99%) were tenth-graders, and 85 (27.07%) were eleventh-graders. Since there were many items left unanswered or marked with more than one answer, or where the same answer was marked for all questions, we get the impression that some students fulfilled the scale without reading the items. As such, we eliminated their responses from the dataset (14 students in total: six girls and eight boys). Then, we obtained the number of students used for the first research group. The second group comprised 266 students, 129 (48.50%) of whom were female and 137 (51.50%) of whom were male, studying at two different high schools in Diyarbakır city center. Of these students, 81 (30.45%) were ninth-graders, 97 (36.47%) were tenth-graders, and 88 (33.08%) were eleventh-graders. Some of the students did not respond the items in a proper way (12 students in total: five girls and seven boys) their data were incomplete or falsely recorded. We eliminated their information from the dataset, and we then obtained the number of students used for the second research group. In order to test the construct validity of the CAES's measures, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were performed on the data collected from the first and second study groups, respectively. Also, Cronbach's alpha was used to determine the reliability of the measures, and item analyses were performed on the data obtained by combining the first and second study groups (consisting of 580 participants). The third study group consisted of 144 students, of whom 83 (57.60%) were girls and the remaining 61 (42.40%) were boys; all of these students attended a high school in the central district of Diyarbakir. Of these students; 27 (18.80%) were ninth-graders, 34 (23.60%) were tenth-graders, and 83 (57.60%) were eleventh-graders. As part of the concurrent validity study, the correlation between students' learning orientations and their perceptions of the classroom assessment environment was calculated using the data obtained from this group. Finally, the fourth study group consisted of 85 students, of whom 29 (34.12%) were girls and 41 (48.24%) were boys; they all attended a high school in the central district of Diyarbakir. Of these students, 29 (34.12%) were ninth-graders, 28 (32.94%) were tenth-graders, and 28 (32.94%) were eleventh-graders. The test–retest reliability of the measures was assessed using the data obtained from this group. Before performing the statistical analyses related to test–retest reliability, data produced by nine students (five girls and four boys) who had not attended either one of the two applications performed three weeks apart from each other were removed from the dataset. Then, the test–retest reliability of the measures was assessed based on the data obtained from the remaining 76 students (39 girls and 37 boys).

Various factors were influential in choosing our research group. Taking into consideration the accessibility principle, we conducted our research in Batman and Diyarbakir. Our research was conducted in seven high schools; of them, three were in Batman (one science high school, one Anatolian high school, one Anatolian imam hatip high school) and four were in Diyarbakir (one science high school and three Anatolian high schools). It needs to be emphasized that these high schools are owned by the Turkish government. Students who participated in the research were in grades 9 to 11. Students in grade 12 had to study for the university entrance exam; they received rapport and did not attend school in March 2014. Since we were collecting data during that period, we could not communicate with them. As a result, our research group did not include students in grade 12. Lastly, it needs to be mentioned that there are various factors that must be taken into account given that we conducted our research with four different groups. The literature suggests that the sample CFA and the sample EFA need to be different from each other (Fabrigar, Wegener, MacCallum & Strahan, 1999). Accordingly, when applying EFA and CFA, we collected data from different groups. During the concurrent validity study, it needs to administer at the same time the CAES and another scale in order to find the correlation between the two scales. Hence, items on the data collection tool increased, and data collection became more difficult when compared to the data collection process conducted during the EFA and CFA. We thus conducted a study with a third group apart from the EFA and CFA. Since it was difficult to administer the CAES to the same students twice, studies of test–retest reliability were conducted with a fourth group, which consisted of few participants. Table 1 presents the scales applied to the research groups and the statistical operations used on the data collected from each group.

Table 1. Study Groups Included in the Research

Study Groups	Employed Scale/Scales	Performed Statistical Procedures	
The First Study Group	CAES	Performing EFA for construct validity	Additionally, Cronbach's alpha reliability coefficient and items analysis were calculated for the measures of the dataset that combined the first and second study groups
The Second Study Group	CAES	Performing CFA for construct validity and calculating composite reliability	
The Third Study Group	CAES and Learning Orientation Subscale	Correlation calculation between the two scales scores for concurrent validity	
The Fourth Study Group	CAES	Correlation calculation between the first and second applications for test–retest reliability	

Data Collection Instrument

When the literature on classroom assessment environment is reviewed, it is evident that the students' perceptions of the classroom assessment environment have a significant effect on their learning orientations. Studies from the literature show that learning orientation is positively correlated with learning-oriented perceptions of the classroom assessment environment, and they are negatively correlated with performance-oriented perceptions of the classroom assessment environment (Alkharusi, 2009). Therefore, when examining the concurrent validity of the comments from the CAES's measures, the correlation between students' perceptions of the classroom assessment environment and their learning orientations was calculated. In order to measure the students' learning orientations, the learning orientation subscale of the Achievement Goal Orientations Scale (AGOS) was employed. The AGOS was developed by Midgley et al. (1998) and adapted into Turkish by Akin and Cetin (2007). The scale contains three dimensions: learning orientation, performance approach orientation, and performance avoidance orientation. The AGOS does not produce a total score; the analyses can be performed only on scores received from its subscales. Therefore, the AGOS's subscales can be administered separately. Since it was aimed in this research to determine students' learning orientations, the performance approach and the performance avoidance subscales were not included in the data collection instrument, and only the learning orientation subscale was employed. The six items that reflect learning orientation were administered to the students using 5-point Likert-type grading, as this is how the scale is structured in its original and Turkish versions. In the adaptation study carried out by Akin and Cetin (2007), it was found that the items' factor loads ranged between .47 and .78, whereas the measures' Cronbach's alpha coefficient was found to be .77 and its test–retest reliability coefficient was found to be .95. Since the learning orientation subscale was employed in this research separately from the other two subscales, the factor analysis and reliability analysis were repeated. Findings obtained from the EFA and CFA demonstrated that the factor loads of the learning orientation scales were above the lower limit of .30 (Buyukozturk, 2010; Pallant, 2005). According to the EFA results, items on the learning orientation subscale have factor loads between .40 and .76. On the other hand, according to the CFA findings, the factor loadings of the items in the learning orientation subscale vary between .35 and .75. Moreover, results from the CFA also showed that the fit indexes obtained for the learning orientation scale were acceptable. In the present research, the reliability of the measures obtained using the learning orientation subscale was calculated through Cronbach's alpha method, and its coefficient was found to be .66. Measures with reliability coefficients greater than .70 are considered reliable (Tezbasaran, 1997). On the other hand, a reliability coefficient greater than .60 is considered adequate for reliability in scales that have a limited number of items (Yurtkoru, Sipahi & Cinko, 2010). Therefore, it could be stated that the Cronbach's alpha coefficient calculated for the measures obtained using the learning orientation subscale fall within acceptable limits. Table 2 demonstrates the findings related to the learning orientation scale's factor loads and the reliability of the measures obtained using this subscale.

Table 2. Factor Loads of the Learning Orientation Scale and Reliability of Measures

Learning Orientation Scale Items	Factor Loads		Cronbach's Alpha
	EFA	CFA	
1. I like school work that I'll learn from, even if I make a lot of mistakes.	.57	.38	.66
2. An important reason why I do my school work is because I like to learn new things.	.76	.75	
3. I like school work best when it really makes me think.	.67	.52	
4. An important reason why I do my work in school is because I want to get better at it.	.73	.69	
5. I do my school work because I'm interested in it.	.49	.30	
6. An important reason I do my school work is because I enjoy it.	.40	.31	

(χ^2/sd=1.88, RMSEA=.078, NFI=.89, NNFI=.91, CFI=.95, IFI=.95, SRMR=.060, GFI=.96, AGFI=.91)

Procedure

When developing CAES, the LOAE and POAE dimensions proposed by Alkharusi (2009, 2011) were taken as the basis. The Classroom Assessment Environment Scale developed by Alkharusi (2009, 2011), and its related literature (Ames, 1992; McMillan & Workman, 1998), were used when writing down the items that were going to be included in the CAES. We emailed Alkharusi to obtain permission to use his scale. Then, some of the items on the scale, as developed by Alkharusi (2009, 2011), were translated to Turkish and added to the CAES item pool. For instance, the item "*In this class, the assignments and activities are related to students' everyday lives*" was translated verbatim to Turkish. Yet, some items that featured cultural and linguistic differences were partly changed during translation, and they were then added to the CAES item pool. To establish an item pool that completely reflected the structure to be measured; we consulted two experts on scale items preparation. One of these experts studied measurement and evaluation, and the other expert studied curriculum and instruction. Both of them obtained their PhD in their field of study. Following their suggestions, the item "*In this class, exam results intend to inform students about their academic development*" was added to the scale. The CAES's item pool was produced based on the LOAE and POAE dimensions. An item pool was formed, with nine items that reflected the LOAE and ten items that reflected the POAE. For the items in the scale, a 5-point Likert-type grading scale was used: *Strongly Agree* (5), *Agree* (4), *Unsure* (3), *Disagree* (2), and *Strongly Disagree* (1). Then, in order to ensure the CAES's content and face validity, opinions were taken from three assessment and evaluation experts and from two curriculum and instruction experts. Given that defining the content (scope) related to a subject requires judgment, experts and developers of the assessment instrument should share definitions of the concept being studied (Tavsancil, 2010). Especially in multidimensional assessment instruments that consist of more than one sub-scales, experts should decide whether the items, written down to determine the different dimensions of the construct that is aimed to be measured, are relevant (or not) to the dimension that they are expected to fall under (DeVellis, 2003). Thus, the experts were asked to assess the scale based on the LOAE and POAE dimensions, which had been taken as the basis for scale development during the initial process. Based on the experts' opinions, it was decided that no items needed to be removed from the LOAE. On the other hand, of the experts we consulted, 3 of them said that instead of the item, "*It is hard to have good grades in the exams of this class*", it would be better to use the item, "*The exams of this class are hard for students*", so as to evaluate POAE. Following their suggestion, we eliminated this item from the scale. Then, 18 items (nine reflecting the LOAE and nine reflecting the POAE) remained on the scale. Since the scale included other items that could measure the feature that the removed item was supposed to measure, the scale's content validity was not harmed. Moreover, based on the experts' opinions, some items were rephrased. For instance, two of the five experts suggested a revision to the following item in order to underline that the criticism is directed towards the students: "*In this classroom, a student's lower performance compared to other students sometimes leads to criticism*." Therefore, this item was rephrased as follows: "*In this classroom, a student's lower performance compared to other students sometimes causes him/her to be criticized*." After making the necessary revisions to the scale's items based on the experts' opinions, the opinions of two experts working for the Turkish Language Society were sought in order to ensure the comprehensibility of the instrument. Based on these experts' opinions regarding spelling and punctuation rules, the scale items were reviewed. Then, in order to receive feedback regarding the comprehensibility of the CAES's items and the duration of the scale's application, a preliminary administration was carried out with 20 high school students (9 girls and 11 boys). These students were then interviewed in order to get their opinions about the comprehensibility of the items. They were also asked for their opinions about the instructions presented at the beginning of the scale, which were meant to inform the participants about the scale's purpose, the number of items used, and how it should be filled out. In the interviews, it was determined that students had difficulty in understanding the following POAE item: "*In this classroom, the grading system used for exams is not clear*." Therefore, this item was rephrased as follows: "*In this classroom, the grading system used for exams is not meant to show what is expected from students*." Students' opinions suggested that no revisions were necessary for the instructions section. The average length of time to fill out the scale was determined by averaging the amount of time spent by the fastest and the slowest students in the preliminary group. The preliminary group's data were not added to the principal application. Once all of these procedures were completed, the scale was ready for implementation to a larger study group.

Data Analysis

After the CAES was administered to the study group, statistical analyses were performed in order to reveal the psychometric properties of the measures. First, the EFA was performed for the construct validity of the CAES's measures. Before performing the EFA, it was necessary to determine whether the dataset was suitable for factor analysis or not. Sample size is the first factor that needs to be considered. There are different opinions regarding the number of participants that should be included in factor analysis studies. Cattell (1978) maintains that in factor analysis studies, the number of participants should be 3–6 times greater than the number of items, and that 200 participants is acceptable while 500 participants is considered to be a highly sufficient number. Hair, Anderson, Tatham and Grablowsky (1979) recommend that the number of participants should be 20 times as many as the number of scale items in the factor analysis. Comrey and Lee (1992) suggest that for the factor analysis, 100 participants is an insufficient number, 200 is mediocre, 300 is good, 500 is very good, and 1000 is perfect (Akbulut, 2010). Furthermore, Ferguson and Cox (1993) state that 100 participants should be the minimum for the factor analysis. As for Kline (1994), he believes that 200 is generally satisfactory to obtain reliable results from the factor analysis, but in cases where the factor structure is clear and small, this can be reduced to 100; however, working with large samples is more appropriate. In estimates of appropriate sample sizes for use in factor analysis, meeting at least two of the criteria available in the literature is recommended (Cokluk, Sekercioglu & Buyukozturk, 2012). In the present study, data from 314 participants from the first study group were used for the EFA. According to these criteria, it can be stated that the number of participants in this study is sufficient for factor analysis. Another operation that needs to be performed for the same purpose is the examination of the Kaiser–Mayer–Olkin (KMO) value, as well as Bartlett's test. The data are deemed suitable for factor analysis if the KMO values are above .60 and if Bartlett's test is statistically significant (Buyukozturk, 2010).

The EFA includes a variety of factorization techniques such as principal components analysis, maximum likelihood factor analysis, image factor analysis, and unweighted least-squares analysis (Tabachnick & Fidell, 2007). Stevens (1996) claimed that the principle components analysis, when compared to many factorization techniques, is psychometrically more powerful and mathematically easier to perform; it is also more efficient in solving the problem of potentially uncertain factors. For that reason, he stated the principal components analysis needs to be a priority when selecting the factorization technique (Steven's study cited in Akbulut, 2010). Taking these arguments into consideration, we used the principal components analysis in our research. The fact that the correlation coefficient between the CAES's dimensions was statistically significant in the study carried out by Alkharusi (2009) led us to believe that the scale's factors would be correlated. Due to this estimation, the direct oblimin rotation in EFA was used. In our evaluation of the EFA results, we followed the following rule: in order for an item to be included on the scale, its factor load – at its theoretically expected dimension – needs to be higher than .30 (Buyukozturk, 2010; Pallant, 2005). With regard to the EFA findings, in addition to the item factor loads, the common measures of variance (h^2) for the variable that will be measured also need to be taken into consideration. In the factor analysis, it is recommended that items with low common variance should be excluded from the instrument (Kalayci, 2010). When interpreting common variance values, it is generally agreed upon that the value of .50 should be taken as a criterion (Thompson, 2004). However, in the social sciences, it is sometimes not possible to obtain high common variance values. Therefore, Costello and Osborne (2005) argue that taking the value of .40 as a criterion is a better choice. As for Tabachnick and Fidell (2001), they explain that items with a common variance of .20 or lower point to heterogeneity among the items (Cokluk, Sekercioglu & Buyukozturk, 2012). From this view, the criterion related to common factor variance should be set at .20 (Sencan, 2005).

We used the CFA to check the EFA results and the measurement model that was theoretically constructed. When the value χ^2 is significant, it is generally accepted that the data do not support the theoretical model. However, the value χ^2 is highly sensitive to sample size. While it has no practical value, the χ^2 value can be statistically significant in large research samples (Byrne, 2010; Kline, 2011). Therefore, the standardized value of χ^2 and the other fit indexes need to be considered first, and then one must decide whether the research data justify the theoretical model (Hu & Bentler, 1999). Numerous fit indexes are used to demonstrate the adequacy of the model tested in the CFA. In this study, the following indexes were examined for CFA: chi-square goodness of fit test, goodness of fit index (GFI), adjusted goodness of fit index (AGFI), comparative fit index (CFI), normed fit index (NFI), non-normed fit index (NNFI), incremental fit index (IFI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), parsimony normed fit index (PNFI), and parsimony goodness of fit index (PGFI). Determination of the criteria to be considered for the fit indexes is a controversial topic (Wetson & Gore, 2006). Nevertheless, in general, values lower than 2 for χ^2/sd show a perfect fit, whereas values between 2 and 3 point to an acceptable fit (Kline, 2011; Schermelleh-Engel, Moosbrugger & Müller 2003). For GFI, CFI, NFI, NNFI, and IFI: .90 points to an acceptable fit, whereas .95 points to a perfect fit (Bentler, 1980; Bentler & Bonett, 1980; Marsh, Hau, Artelt, Baumert & Peschar, 2006). For AGFI: .85 shows an acceptable fit and .90 shows a perfect fit (Schermelleh-Engel & Moosbrugger, 2003). For RMSEA: .08 points to an acceptable fit and .05 points to a perfect fit (Brown & Cudeck, 1993; Byrne & Campbell, 1999). For SRMR: .10 means that the fit is acceptable and .05 means that it is a perfect fit (Hu & Bentler, 1999; Kline, 2011). For PNFI and PGFI: values above .50 point to an acceptable fit (Meyers, Gamst & Guarino, 2006), whereas values equal to or above .95 point to a perfect fit (Meydan & Sesen, 2011).

As part of the concurrent validity, the correlation between students' CAES scores and their learning orientations was calculated. The reliability of the measures obtained using the CAES was calculated through Cronbach's alpha coefficient, composite reliability, and the test–retest method. In order to determine the CAES's items' discriminatory power, the adjusted item total correlation was examined, and 27% sub-upper group comparisons were made. SPSS 20.0 software was employed for the analyses of the EFA, concurrent validity, Cronbach's alpha and the test–retest reliability, and the item analyses. On the other hand, calculations pertaining to the CFA were performed using LISREL 8.54 software. Finally, the composite reliability coefficient of the measures was calculated manually by computing formulas given by Fornell and Larckers (1981) using the Microsoft Excel 2010.

$$p_c = \frac{\left(\sum_{i=1}^{m} \lambda_i\right)^2}{\left(\sum_{i=1}^{m} \lambda_i\right)^2 + \left(\sum_{i=1}^{m} (\theta_i)\right)}$$

p_c = Composite Reliability Coefficient

λ_i = Standardized Factor Loadings

θ_i = Measurement Error of Item (Standard Error Variance)

Results

Construct Validity

The EFA and CFA were performed to examine the construct validity of the CAES's measures.

Explanatory Factor Analysis (EFA)

In this study, the KMO value was found to be .771 and Bartlett's test was found to be statistically significant (χ^2=883.450, *sd*=153). Therefore, it could be concluded that our data were suitable for the factor analysis. Following this finding, and as a result of the EFA principle components factorization technique and direct oblimin rotation (delta=0, kappa=4), a construct with two factors that explained 31.52% of the total variance was obtained. Table 3 presents the findings obtained in the EFA.

Table 3. CAES's Factor Structure and Factor Loads

Factor	Item No	Statements	Factor Load		h^2
			Factor 1	Factor 2	
Factor 1: LOAE	I1	Assessment practices performed in this classroom enable students to see their strengths and weaknesses	.54	-.12	.30
	I3	Assessment practices performed in this classroom help students to decide what subjects they need to study more.	.62	-.14	.38
	I5	Feedback is given in this classroom to students to enable them to iron out their performance-related problems.	.33	-.22	.15
	I7	Opportunities are given in this classroom for students to correct their learning-related mistakes.	.43	-.36	.26
	I9	Assessment activities performed in this classroom are linked to students' daily lives.	.47	.04	.24
	I11	In this classroom, students are also given responsibility in the assessment process.	.52	-.12	.27
	I13	In order to determine students' levels of comprehension on a certain subject, various assessment methods are used in this classroom, which include some of the following tasks: class discussions, project assignments, multiple choice tests, homework, and written and verbal exams.	.55	-.04	.31
	I15	In this classroom, exam results are meant to provide information about the student's progress.	.57	-.02	.33
	I17	Assessment techniques employed in this classroom lead students to think.	.63	-.05	.40
		% Variance Explained	**20.926**		
Factor 2: POAE	I2	Exams are difficult for students in this classroom.	-.16	.56	.32
	I4	In this classroom, teachers attach more importance to students' grades than to whether students comprehend the subject or not.	-.19	.65	.42
	I6	Assignments and exams are not in concordance with one another in this classroom.	-.44	.52	.39
	I8	The grading system employed for exams in this classroom is not meant to show what is expected from students.	-.45	.33	.26
	I10	In assessments in this classroom, greater emphasis is placed on students' exam grades than on their efforts.	-.21	.69	.48
	I12	The criterion of success in this classroom is whether a student performs better than others.	.03	.44	.21
	I14	In this classroom, subjects covered during classes and assignments given to students are incongruent.	-.54	.39	.37
	I16	In this classroom, a student's lower performance compared to other students sometimes causes him/her to be criticized.	-.15	.58	.34
	I18	Assessment methods used in this classroom are oriented more towards the outputs that they produce than towards the students' efforts.	.02	.50	.27
		% Variance Explained		**10.597**	
		% TOTAL VARIANCE EXPLAINED	**31.524**		

According to the findings in Table 3, it was determined that all of the items in the scale had factor loadings greater than the lower limit of .30. Also, the scale items, excluding item 5 on the CAES, fulfilled the criterion of .20 for common measures of variance. Yet, item 5 was important in regard to scale's content validity, and it exhibited an adequate factor load in both the EFA and CFA. Thus, we did not eliminate this item from the scale. By considering the contents of the items that were aggregated under the factors, as well as the theoretical structure, the first factor was named LOAE, while the second was named POAE. The LOAE consists of nine items, which explained 20.93% of the total variance. Factor loads of the LOAE's items ranged between .33 and .63. On the other hand, the POAE consisted of nine items, which contributed 10.60% of the total explained variance. Factor loads of the POAE's items ranged between .33 and .69. Since the factor loads of all the items in the scale were above the lower limit of .30 (Costello & Osborne, 2005), there was no need to remove any item from the scale. According to the EFA findings, items 8 and 14 – which were theoretically expected to be placed in POAE – had a higher factor load in LOAE. However, both items had an adequate factor load in POAE. We thus decided to place these items in POAE. The CFA results concerning the CAES also support this view.

Confirmatory Factor Analysis (CFA)

We applied CFA to assess whether the second group's data would justify the EFA results obtained from 18 items and two factors. Fit indexes of the obtained model were examined in the CFA performed, and it was observed that the CAES's fit index values were as follows: χ^2/sd=1.84, GFI=.91, AGFI=.88, CFI=.96, NFI=.91, NNFI=95, IFI=.96, RMSEA=.056, SRMR=.056, PNFI=.79, and PGFI=.71. Table 4 demonstrates the acceptable and perfect fit values, which were examined in order to test the adequacy of the model; also shown are the fit index values obtained from the CFA, as well as the relevant results. The criteria used for acceptable and perfect fit (as demonstrated in Table 4) indicate that the model with two factors obtained from the CFA fits.

Table 4. Acceptable and Perfect Fit Values for Fit Indexes and Fit Index Values obtained in CFA

Fit Indexes Examined	Criteria for Perfect Fit	Criteria for Acceptable Fit	Fit Indexes Obtained	Result
χ^2/sd	$0 \leq \chi^2/sd \leq 2$	$2 \leq \chi^2/sd \leq 4$	1.84	Perfect Fit
GFI	.95 ≤ GFI ≤ 1.00	.90 ≤ GFI ≤ 95	.91	Acceptable Fit
AGFI	.90 ≤ AGFI ≤ 1.00	.85 ≤ AGFI ≤ .90	.88	Acceptable Fit
CFI	.95 ≤ CFI ≤ 1.00	.90 ≤ CFI ≤ .95	.91	Acceptable Fit
NFI	.95 ≤ CFI ≤ 1.00	.90 ≤ CFI ≤ .95	.91	Acceptable Fit
NNFI	.95 ≤ NNFI ≤ 1.00	.90 ≤ NNFI ≤ .95	.95	Perfect Fit
IFI	.95 ≤ IFI ≤ 1.00	.90 ≤ IFI ≤ .95	.96	Perfect Fit
RMSEA	.00 ≤ RMSEA ≤ .05	.05 ≤ RMSEA ≤ .08	.056	Acceptable Fit
SRMR	.00 ≤ SRMR ≤ .05	.05 ≤ SRMR ≤ .10	.056	Acceptable Fit
PNFI	.95 ≤ PNFI ≤ 1.00	.50 ≤ PNFI ≤ .95	.79	Acceptable Fit
PGFI	.95 ≤ PGFI ≤ 1.00	.50 ≤ PGFI ≤ .95	.71	Acceptable Fit

χ^2=246.45, *sd*=134, 90% Confidence interval for RMSEA = (.045, .067)

Table 5 demonstrates the *t*-test values for the model. Table 5 shows that the *t*-test values for the LOAE ranged between 6.31 and 10.24, whereas for the POAE, it ranged between 4.25 and 11.59. Moreover, *t*-values greater than 1.96 point to a level of significance at .05, whereas *t*-values greater than 2.58 point to a significance level of .01 (Jöreskog & Sörbom, 1993; Kline, 2011). Thus, it was concluded that all *t*-values obtained from the CFA were significant at the level of .01. Insignificant *t*-values implied that items related to these *t*-values should be removed from the model, or that the number of participants in the research was inadequate for the factor analysis (Byrne, 2010). Therefore, the *t*-values obtained from the CFA prove that the number of participants in this research is adequate for the factor analysis and that no item should be removed from the model.

Table 5. The *t*-test Values Obtained from the CFA for CAES

Item No	*t*	Item No	*t*	Item No	*t*
LOAE1	8.07*	LOAE7	6.49*	POAE4	8.20*
LOAE2	8.91*	LOAE8	8.38*	POAE5	10.91*
LOAE3	6.31*	LOAE9	10.24*	POAE6	4.83*
LOAE4	9.02*	POAE1	7.41*	POAE7	8.75*
LOAE5	7.21*	POAE2	9.10*	POAE8	9.00*
LOAE6	7.11*	POAE3	11.59*	POAE9	4.25*

*p<.01

Factor loads for the model with two factors obtained during the CFA are presented in Figure 1. As is seen in the figure, factor loads range between .41 and .62 for LOAE and between .28 and .68 for POAE.

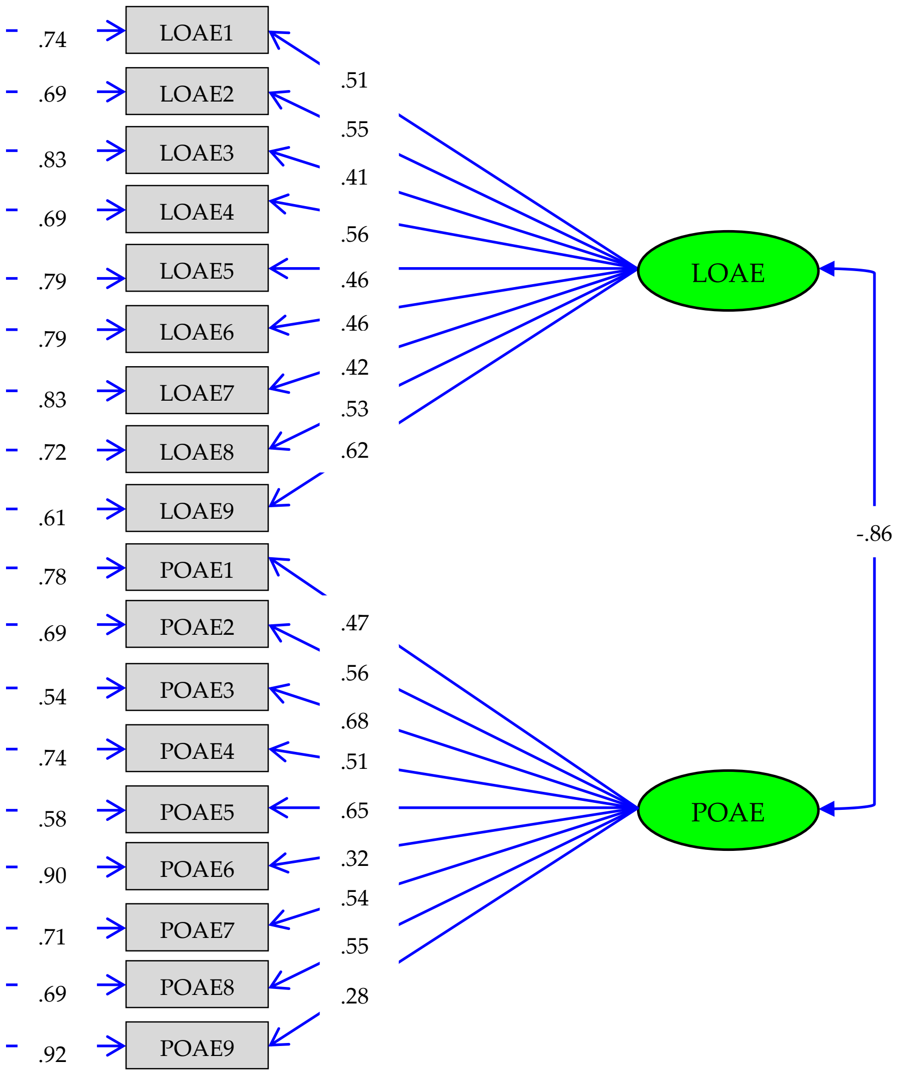

Figure 1. Measurement Model for CAES

Concurrent Validity

Within the scope of concurrent validity, the correlation between students' perceptions of the classroom assessment environment and their learning orientations was calculated. As a result of the correlation analysis, it was found that learning orientation is positively correlated with the LOAE [r=.31, p<.01] and negatively correlated with the POAE [r=-.27, p<.01]. Given the fact that students with learning-oriented perceptions of the classroom assessment environment have higher whereas those of students with performance-oriented perceptions are lower (Wang & Cheng, 2010), the correlations between CAES and learning orientation can be evaluated as a proof for the concurrent validity of the comments based on the CAES's measures.

Reliability

The reliability of the measures obtained from the CAES was calculated using Cronbach's alpha, composite reliability, and the test–retest method. The Cronbach's alpha coefficients were calculated as .73 for measures obtained from both the LOAE and POAE. The composite reliability coefficients were found to be .75 and .76 for measures obtained from the LOAE and POAE subscales, respectively. In order to determine the test–retest reliability of the measures, two applications were conducted with 76 students three weeks apart from each other. For the purpose of revealing the consistency between the first and second applications, correlation coefficients between the scores obtained from the two applications were calculated. The test–retest reliability coefficients were found to be .93 for the measures obtained from both subscales. Given the fact that measures with reliability coefficients greater than .70 are considered reliable (Domino & Domino, 2006; Fraenkel, Wallen, & Hyun, 2012), it could be stated that these reliability coefficients were adequate. Table 6 shows the results of the reliability analysis.

Table 6. The Reliability Coefficients Calculated via Cronbach's Alpha, Composite Reliability, and the Test–Retest Method for Measures Obtained from the CAES's Dimensions

Subscales	Cronbach's Alpha	Composite Reliability	Test–Retest
LOAE	.734	.753	.932
POAE	.730	.761	.927

Item Analysis

With the purpose of determining the items' discriminatory power, as well as their power for predicting the total score, the adjusted item total correlation was examined, and 27% sub-upper group comparisons were made. Table 7 demonstrates the findings obtained through the item analysis.

Table 7. CAES Item Analysis Results

Item No	Scale Alpha If the Item is Deleted	Adjusted Item Total Correlation	Mean	Std Deviation	Skewness	*t*	
LOAE1	.708	.426	3.15	1.25	-.29	16.04*	*sd*=274 *p<.001
LOAE2	.703	.454	3.09	1.28	-.30	15.08*	
LOAE3	.727	.316	2.74	1.31	.00	13.63*	
LOAE4	.714	.389	2.97	1.29	-.15	13.96*	
LOAE5	.716	.380	2.31	1.26	-.56	13.71*	
LOAE6	.711	.410	3.37	1.27	-.59	14.56*	
LOAE7	.717	.375	3.23	1.36	-.41	16.27*	
LOAE8	.708	.426	3.15	1.37	-.29	15.89*	
LOAE9	.693	.518	3.06	1.23	-.28	19.38*	
POAE1	.707	.405	3.63	1.26	-.50	14.05*	*sd*=304 *p<.001
POAE2	.700	.439	3.04	1.52	.00	18.17*	
POAE3	.689	.499	2.98	1.39	.11	19.40*	
POAE4	.714	.361	3.05	1.32	-.07	13.26*	
POAE5	.682	.537	3.47	1.41	-.38	23.78*	
POAE6	.729	.258	3.54	1.19	-.46	8.60*	
POAE7	.705	.416	2.44	1.36	.67	13.98*	
POAE8	.701	.435	3.32	1.41	-.30	16.71*	
POAE9	.727	.283	3.35	1.30	-.28	10.89*	

Table 7 shows that the *t*-values for the differences between item values of the 27% sub- and upper-groups range between 13.63 and 19.38 (*sd*=274, *p*<.001) for the LOAE, and between 8.60 and 23.78 (*sd*=304, *p*<.001) for the POAE. Moreover, according to Table 7, values pertaining to the item total correlation range between .32 and .52 for the LOAE and between .26 and .54 for the POAE. When interpreting item total correlation values, items with values equal to or greater than .30 are considered adequate when rendering the feature to be measured as distinct (Buyukozturk, 2010; Erkus, 2012). All items except for POAE6 and POAE9 met this criterion. On the other hand, it was apparent that the *t*-values for these items were significant. Significant *t*-values related to the differences between the sub- and upper-groups are considered to prove the item's discriminatory ability (Erkus, 2012). Therefore, it could be concluded that all items in the scale are discriminatory.

Interpretation of CAES Scores

The CAES consists of 18 items. On the scale, a 5-point Likert-type grading scale [*Strongly Agree* (5)→*Strongly Disagree* (1)] was used. The scale has two dimensions: the LOAE and POAE, both of which are composed of nine items. Therefore, possible scores for each dimension range between 9 and 45. When assessing the CAES scores, the assessment was performed based on the scores taken from the subscales. That is, a total score related to the classroom assessment environment cannot be obtained on the scale. Higher scores from the subscales indicate that students' perceptions of the relevant dimension are high.

Discussion

The purpose of this research was to develop an assessment instrument that could validly and reliably measure students' perceptions of the classroom assessment environment. When developing the CAES, the dimensions of LOAE and POAE (Alkharusi, 2009; 2011) were used as the base. An item pool was formed with nine items that reflected the LOAE and ten items that reflected the POAE. Experts' opinions were obtained in order to ensure the scale's content and face validity. Based on these opinions, one item was removed from the POAE dimension. The instrument was administered to the students with a 5-point Likert-type grading scale, with responses varying from *Strongly Agree* (5) → *Strongly Disagree* (1).

For construct validity of the measures obtained from the CAES, EFA and CFA were performed. During the EFA, a construct with two factors and 18 items that explained 31.52% of the total variance was obtained. By considering the contents of the items that were aggregated under the factors, as well as the theoretical structure, the first factor was named LOAE while the second was named POAE. CFA was performed in order to determine whether a theoretically constructed measurement model demonstrates satisfactory goodness of fit indices. Findings obtained from the CFA demonstrated that the fit indexes belonging to the construct with two factors were adequate. Given the fact that 30% was taken as the minimum requirement for the ratio of variance explained in the EFA (Bayram, 2009; Buyukozturk, 2010), that the factor loads of all items in the scale were above the lower limit of .30 (Costello & Osborne, 2005), and that the fit indexes obtained from the CFA fell within the acceptable limits, it could be concluded that the construct validity of the CAES's measures was ensured.

The correlations between the CAES's sub-dimensions and learning orientation subscale were calculated so that the concurrent validity could be identified. Findings obtained from the correlation analysis indicated that the students' learning orientations were positively correlated with the LOAE and negatively correlated with the POAE. This finding is supported by other studies in which the relationship between learning orientation and the classroom assessment environment has been addressed (Alkharusi, 2009; Wang & Cheng, 2010). Therefore, it could be concluded that the concurrent validity of the comments from CAES's measures was proven.

The reliability of the measures obtained from the CAES was calculated via Cronbach's alpha, composite reliability, and the test–retest method. The Cronbach's alpha coefficients were calculated as .73 for measures obtained from both the LOAE and POAE. The composite reliability coefficient was .75 for measures obtained from the LOAE subscale and .76 for measures obtained from the POAE subscale. On the other hand, the reliability coefficients obtained through the test–retest method were the same (.93) for the measures obtained from both subscales. Measures with reliability coefficients equal to or greater than .70 were considered reliable (Fornell & Larcker, 1981; Tezbasaran, 1997; Nunnaly & Bernstein, 1994). Therefore, the reliability coefficients obtained through Cronbach's alpha, composite reliability, and the test–retest method can be regarded as proof for the reliability of the measures. The finding that the reliability coefficients obtained through the test–retest method are higher than those obtained through Cronbach's alpha and composite reliability means that the reliability of the measures – in terms of their stability – is higher than its reliability in terms of their consistency.

An item analysis was carried out in order to determine the how well the CAES's items predicted the total score, as well as the items' levels of distinctiveness. As part of item analysis, the adjusted item total correlation was examined and 27% sub-upper group comparisons were made. The adjusted item total correlations were found to be between .32 and .52 for the LOAE and between .26 and .54 for the POAE. It was also found that the *t*-values obtained in the sub-upper group comparisons were significant for all items. These findings suggest that all of the CAES's items are discriminatory. The findings obtained in this research via statistical analyses, with the ultimate aim of examining the psychometric properties of the measures presented in the CAES, demonstrate that the scale can be used an instrument that produces valid and reliable measures to determine students' perceptions of the classroom assessment environment.

Conclusions

The literature shows that there a non-Turkish scale exists that can be used to measure students' perceptions of the classroom assessment environment. However, no assessment instrument that could serve the same purpose has been found in the relevant literature in Turkish. Therefore, it is believed that this study – the aim of which was to develop the CAES – will significantly contribute to the relevant literature. In other words, one of the strengths of this research was that we were able to carry out a study on the classroom assessment environment that could be used in the Turkish literature for the first time. The presentation of more than one evidence related to the construct validity, reliability, and item discrimination of the CAES is another strength aspect of the present study. For instance, in the reliability analysis of the measures, Cronbach's alpha, composite reliability, and test–retest reliability coefficients were used. The reliability of the measures in terms of their consistency was investigated by means of Cronbach's alpha and composite reliability. On the other hand, the reliability of the measures in terms of their stability (invariance against time) was examined via the test–retest method. In order to demonstrate the discriminatory power of the scale items, the corrected item total correlation was performed and the lower and upper groups in the 27% range were compared. With regard to construct validity of the CAES's measures, both EFA and CFA were performed.

Limitations of the Research and Suggestions for Future Research

In addition to the strengths aspects listed above, the present study has some limitations. These limitations bring about a number of suggestions for further research. First, this scale development study was carried out only with high school students. It is known that reliability and validity are the characteristics of the measures obtained from the scale and the comments based on these measures respectively (Bademci, 2013). For this reason, studies should be carried out with different samples, as this is important for the replication of the validity and reliability analyses. Another suggestion for further research includes analyzing the concurrent validity of the CAES's measures. In this study, within the scope of concurrent validity, the relationship between the students' LOAE and POAE scores and their learning orientations was examined. The relevant literature suggests that the classroom assessment environment plays a significant role in establishing students' beliefs about their self-efficacy (Alkharusi, 2009), as well as their motivational beliefs and academic achievements (Alkharusi, 2007). Thus, future studies might investigate the kind of relationship that exists between perceptions of the classroom assessment environment and the aforementioned variables. Using the CAES in future studies is also of importance, as this will contribute to the instrument's power to measure these factors.

References

Akbulut, Y. (2010). *SPSS applications in social studies*. İstanbul: İdeal Publishing.

Akin, A., & Cetin, B. (2007). Achievement goal orientations scale: The study of validity and reliability. *Eurasian Journal of Educational Research, 26*, 1-12.

Alkharusi, H. (2007). *Effects of teachers' assessment practices on ninth grade students' perceptions of classroom assessment environment and achievement goal orıentations in muscat science classrooms in the Sultanate of Oman*. Unpublished Doctoral Dissertation, Kent State University, Oman.

Alkharusi, H. (2008) Effects of classroom assessment practices on students' achievement goals. *Educational Assessment, 13*(4), 243-266.

Alkharusi, H. (2009). *Classroom assessment environment, self-efficacy, and mastery goal orientation: A causal model*. Paper Presented at the 2nd International Conference of Teaching and Learning (ICTL 2009), INTI University College, Malaysia.

Alkharusi, H. (2010a). Teachers' assessment practices and students' perceptions of the classroom assessment environment. *World Journal on Educational Technology, 2*(1), 27-41.

Alkharusi, H. (2010b). A multilevel linear model of teachers' assessment practices and students' perceptions of the classroom assessment environment. *Procedia Social and Behavioral Sciences, 5*, 5-11.

Alkharusi, H. (2011). Development and datametric properties of a scale measuring students' perceptions of the classroom assessment envıronment. *International Journal of Instruction, 4*(1), 1308-1470.

Alkharusi, H. (2013). Canonical correlational models of students' perceptions of assessment tasks, motivational orientations, and learning strategies. *International Journal of Instruction*, *6*(1), 21-38.

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2014). Modeling the relationship between perceptions of assessment tasks and classroom assessment environment as a function of gender. *Asia Pacific Education Researcher*, *23*(1), 93-104.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, *84*(3), 261-271.

Bademci, V. (2013). Estimation of interrater score reliability by the Cronbach's Alpha. *The Journal of the Arts Education Faculty of Gazi University*, *30*, 55-62.

Banks, S. R. (2012). *Classroom assessment: Issues and practices*. IL: Waveland Press Inc.

Bartha, J. M., Dunlap, S. T., Dane, H., Lochmanb, J. E., & Wells, K. C. (2004). Classroom environment influences on aggression, peer relations, and academic focus. *Journal of School Psychology*, *42*(2), 115-133.

Bayram, N. (2009). *Data analysis through SPSS in social sciences*. Bursa: Ezgi Publishing.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, *31*, 419-456.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5-12.

Brookhart, S. M ., & Bronowicz, D. L. (2003). I don't like writing. It makes my fingers hurt: Students talk about their classroom assessments, assessment in education: *Principles, Policy & Practice, 10*(2), 221-242.

Brookhart, S. M., & DeVoge, J. G. (1999) Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education, 12*(4), 409-425.

Brookhart, S. M., & Durkin, D. T. (2004). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education, 16*(1), 27–54.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In: Bollen, K. A., & Long, J. S. (Eds.), *Testing Structural Equation Models* (pp. 136-162). Beverly Hills, CA: Sage

Buyukozturk, S. (2010). *A manual on data analysis in social sciences*. Ankara: Pegem Academy Publishing.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. New York, NY: Taylor and Francis Group.

Byrne, B., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology, 30*(5), 555-574.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum.

Church, M. A., Elliot, A. J., & Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology, 93*(1), 43-54.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7), 1-9.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth: Holt, Rinehart and Winston Inc.

Cronbach, L. J. (1984). *Essentials of psychological testing*. New York: Harper Row.

Cokluk, O., Sekercioglu, G., & Buyukozturk, Ş. (2012). *Multivariate statistics for the social sciences: SPSS and LISREL applications*. Ankara: Pegem Academy Publishing.

Dart, B., Burnett, P. C., Boulton-Lewis, G., Campbell, J., Smith, D., & McCrindle, A. (1999). Classroom learning environments and students' approaches to learning. *Learning Environments Research, 2*(2), 137–56.

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Newbury Park: Sage Publications.

Domino, G., & Domino, M. L. (2006). *Psychological testing: An introduction*. Cambridge: Cambridge University Press.

Dorman, J. P. (2001). Associations between classroom environment and academic efficacy. *Learning Environments Research, 4*(3), 243-257.

Erkus, A. (2012). *Measurement and scale development in psychology*. Ankara: Pegem Academy Publishing.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment, 1*(2), 84-94.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.

Fraenkel, J. R., Wallend, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. New York: McGraw Hill.

Griffin, B. (2009). An examination of student achievement in English language arts through the use of predictive assessments, standardized testing, and academic interventions. *Education Doctoral*. Paper 93.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Grablowsky, B. J. (1979). *Multivariate data analysis*. Tulsa, OK: Pipe Books.

Harbaugh, A. G., & Cavanagh, R. F. (2012, December) *Associations between the classroom learning environment and student engagement in learning 2: A structural equation modeling approach*. Paper Presented at the AARE-APERA 2012, Sydney.

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

Jöreskog, K., & Sörbom, D. (2000). *LISREL* [Computer Software]. Lincolnwood, IL: Scientific Software, Inc.

Kalayci, S. (2010). Faktör analysis. In S. Kalayci, (Eds.), *Multivariate statistical techniques through SPSS* (pp. 234-255). Ankara: Asil Yayın Publishing.

Kline, R. B. (1994). *An easy guide to factor analysis*. New York: Routledge.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Kose, E., & Kucukoglu, A. (2008). Evaluation of class learning environment in faculties of education in terms of some variables. *Ahi Evran University Journal of Kırşehir Education Faculty*, 10(3), 61-73.

Kurt, H., Ekici, G., Gökmen, A., Aktaş, M., & Aksu, Ö. (2013). The effect of learning styles on secondary education students' perception of classroom environment of biology laboratory (Kolb Learning Style Model Sample). *Turkish Studies, 8*(6), 157-177.

Lau, K., & Lee, C.K.J. (2006, November). *Perception of classroom environment, achievement goals, and strategy use among Hong Kong students*. Paper Presented at the Asia-Pacific Educational Research Association Conference 2006, Hong Kong.

Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing, 6*(4), 311-360.

McMillan, J. H., & Workman, D. J. (1998). *Classroom assessment and grading practices: A review of the literature*. Richmond, VA (ERIC Document Reproduction Service No. ED453263): Metropolitan Educational Research Consortium.

Mertler, C. A. (2003, October). *Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the meeting of the Mid-Western Educational Research Association, Columbus, OH.

Mesa, V. (2012). Achievement goal orientations of community college mathematics students and the misalignment of instructor perceptions. *Community College Review, 40*(1), 46-74.

Meydan, C. H., & Sesen, H. (2011). *Structural equation modeling and AMOS applications*. Ankara: Detay Publishing.

Meyers, L. S, Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. London: SAGE Publications.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill.

Pallant, J. (2005). *SPSS survival manual: A step by step guide to data analysis using SPSS for windows*. Australia: Australian Copyright.

Phan, H. P. (2012). Achievement goals, classroom environment, and reflective thinking: A conceptual framework. *Electronic Journal of Research in Educational Psychology, 6*(3), 571-602.

Popilskis, L. B. (2013). *The link between students' achievement goals and perceptions of the classroom environment*. Unpublished Doctoral Dissertation, Fordhman University, New York.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research-Online, 8*(2), 23-74.

Schriesheim, C. A., & Eisenbach, R. J. (1995). An exploratory and confirmatory factor analytic investigation of item wording effects on obtained factor structures of survey questionnaire measures. *Journal of Management, 21*(6), 1177-1193.

Sipahi, B., Yurtkoru, S., & Cinko, M. (2010). *Data analysis through SPSS in social sciences*. Istanbul: Beta Publishing.

Stangor, C. (2010). *Research methods for the behavioral sciences*. Wadsworth.

Stiggins, R. J., & Conklin, N. F. (1992*). In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.

Sencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlilik*. Ankara: Seçkin Yayıncılık.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston, Pearson Education, Inc.

Tavsancil, E. (2009). *Measuring attitudes and data analysis with SPSS*. Ankara: Nobel Publishing.

Tezbasaran, A. (1997). *Guide to develop Likert type scales*. Ankara: Turkish Psychological Association.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington DC: American Psychological Association

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1994). Educational resilience in inner cities. In M. C. Wang & E. Gordon (Eds.), *Educational resilience in inner-city America: Challenges and prospects* (pp. 45-72). Hillsdale, NJ: Erlbaum.

Wang, X., & Cheng, L. (2010). Chinese EFL students' perceptions of classroom assessment environment and their goal orientations. In Cheng, L., & Curtis, A. (Eds.), *English language assessment and the Chinese learner* (pp. 202-218), Taylor & Francis.

Wannarka, R., & Ruhl, K. (2008). Seating arrangements that promote positive academic and behavioural outcomes: A review of empirical research. *Support for Learning*, 23(2), 89-93.

Wetson, R., & Gore Jr, P. A. (2006). A brief guide to structural equation model. *The Counseling Psychologist*, *34*(5), 719-751.

Yuen-Yee, G. C., & Watkins, D. (1994). Classroom environment and approaches to learning: An investigation of the actual and preferred perceptions of Hong Kong secondary school students. *Instructional Science*, *22*(3), 233-246.