Adaptation of Engagement Questionnaire to Turkish for Science Classes: Validity and Reliability Study

Öğrenci Katılımı Ölçeği'nin Fen Bilimleri Dersi için Türkçe'ye Uyarlanması: Geçerlik ve Güvenirlik Çalışması

Fatma Melike Uçar, *Hasan Kalyoncu University, melike.ucar@hku.edu.tr*
Semra Sungur, *Middle East Technical University, ssungur@metu.edu.tr*

Abstract. The aim of this research was to adapt the Engagement Questionnaire in four dimensions (i.e. behavioral, emotional, cognitive, and agentic engagement) to Turkish to be used in science classes. For this purpose, firstly a pilot study was conducted with 153 middle school students, then the main study was conducted with 744 middle school students. Both studies supported four-factor structure of Turkish version of the EQ. In addition, reliabilities seemed to be sufficient. The results concerning the measurement invariance across gender revealed the data from both genders lead to equally valid conclusions regarding their engagement levels. Consequently, the findings of this study indicated that the Turkish version of EQ appears to be a valid and reliable instrument to measure and understand the middle school students' engagement level in science classes.

Keywords: Student Engagement, Validity, Reliability, Confirmatory Factor Analysis, Measurement Invariance across Gender

Öz. Bu araştırmanın amacı, davranışsal, duygusal, bilişsel ve aracı katılım olmak üzere dört alt boyuttan oluşan Katılım Ölçeği'ni fen bilimleri dersinde kullanılmak üzere Türkçe'ye uyarlayarak geçerlik ve güvenirliğini test etmektir. Bu amaçla ölçek, Türkiye'deki ortaokullardan seçilen 153 öğrenciye uygulanarak pilot çalışma yapılmış, ardından 744 öğrenciye uygulanarak ana çalışma yapılmıştır. Her iki çalışmada da yapılan doğrulayıcı faktör analizi sonuçlarında tüm maddeler özgün formunda yer aldıkları faktörlerde toplanmıştır. İç tutarlığın yeterli düzeyde olduğu bulunmuştur. Ayrıca, yapılan ölçme değişmezliği analizi, ölçeğin kızlar ve erkekler için değişmezlik gösterdiğini ortaya koymuştur. Sonuç olarak, bu çalışma Katılım Ölçeği'nin Türkçe formunun geçerli ve güvenilir bir ölçek olarak ortaokul öğrencilerinin fen derslerindeki katılımlarının her bir boyutunu ölçmek için kullanılabileceğini önermektedir.

Anahtar Sözcükler: Öğrenci Katılımı, Geçerlik, Güvenirlik, Doğrulayıcı Faktör Analizi, Ölçme Değişmezliği

ÖZET

Amaç ve Önem

Öğrenci katılımı kavramı, öğrecilerin akademik sağlamlıklarını ve başarılarını şekillendirmesi ve aynı zamanda yüksek düzeydeki sıkılma ve okulu bırakma durumuna ve düşük başarı seviyesine karşı bir çözüm olarak görülmesi nedeniyle son yıllarda oldukça önem kazanmıştır. Ayrıca, öğrenci katılımının öz yeterlik, motivasyon, ve akademik başarı gibi pek çok akademik çıktıyı yordadığı bilinmektedir. Bu nedenle, ülkemizdeki özellikle fen derslerindeki düşük başarı durumu ve fen dersine karşı oluşan olumsuz tutum da göz önünde bulundurulduğunda öğrenci katılımının önemi daha da artmaktadır. Ancak öğrenci katılımını ölçmek için geliştirilen ulusal ve uluslararası ölçekler incelendiğinde bu ölçeklerin çoğunun farklı nicelik ve nitelikteki alt boyutlardan oluştuğu görülmüştür. Dolayısıyla öğrenci katılımını en doğru şekilde değerlendiren, geçerli ve güvenilir bir ölçeğin Türkçe'ye çevrilmesi ve fen derslerine adapte edilerek fen derslerinde öğrencilerin katılım seviyelerinin belirlenebilmesi önem arz etmektedir. Bu nedenle bu çalışmada Reeve ve Tseng tarafından 2011 yılında geliştirilen Katılım Ölçeği'nin Türkçe'ye uyarlanması ile ulusal literatüre bu alanda geçerli ve güvenilir bir ölçeğin sağlanması amaçlanmıştır.

Yöntem

Çalışmanın verileri pilot çalışma için Türkiye'deki ortaokul öğrencilerinden seçilen 153 öğrenciden; ana çalışma için ise 744 öğrenciden toplanmıştır. Verilerin toplanmasında "Öğrenmede Güdüsel Stratejiler Anketi" ve "Katılım Anketi" kullanılmıştır. Katılım Anketi, dört alt boyuttan (davranışsal, duygusal, bilişsel ve aracı katılım) oluşan toplam 22 maddelik 7'li Likert tipinde bir öz-bildirim ölçeğidir. Öğrenmede Güdüsel Stratejiler Anketi'nin bir boyutu öğrencilerin öz-yeterlik seviyelerini belirlemek amacıyla kullanılmıştır. Pilot ve ana çalışma için doğrulayıcı faktör analizi (DFA) ve güvenirlik analizleri yapılmıştır. Ayrıca cinsiyetler arası ölçme değişmezliği incelenmiş ve geçerliğin ileri değerlendirmesi de yapılmıştır.

Bulgular

Pilot çalışma için yapılan DFA sonucu iyi bir model uyumuna işaret etmiştir (CFI = .98, RMSEA = .05, SRMR= .05). Ayrıca açıklayıcı faktör analizindeki faktör yüklerine karşılık gelen tüm Lambda – X tahminleri yeterince yüksek ve istatiksel olarak manidardır. Güvenirlik analizlerinde ise Cronbach alfa katsayıları aracı katılım için .82, davranışsal katılım için .92, duygusal katılım için .84, ve bilişssel katılım için .86 olarak hesaplanmıştır. Ana çalışma için yapılan DFA sonuçları da benzer şekilde iyi bir model uyumu olduğunu göstermiştir (CFI = .99, RMSEA = .05, SRMR = .04). Ayrıca örüntü katsayıları ve yapı katsayıları, tüm maddelerin ilgili oldukları faktörlerle yüksek korelasyon gösterdiği ve aynı zamanda diğer faktörlerle de düşük oranda ilişkili olduğunu göstermiştir. Cronbach alfa değerleri ise aracı katılım için .82, davranışsal katılım için .88, bilişssel katılım için .86 ve duygusal katılım için .83 olarak hesaplanmıştır. Cinsiyetler arası ölçme değişmezliği incelendiğinde ise sonuçlar, faktör yükleri, kovaryanslar, varyanslar ve hata terimlerinin cinsiyetler arasında değişmez olduğuna işaret etmektedir. Geçerliğin ileri değerlendirmesinde ise beklenildiği gibi özyeterliğin manidar olarak aracı katılımla (r = .53), davranışsal katılımla (r = .57), bilişssel katılımla (r = .60) ve duygusal katılımla (r = .58) pozitif yönde ilişkili olduğu görülmüştür.

Tartışma, Sonuç ve Öneriler

Bu çalışma, yapılan analizler sonucunda katılım ölçeğinin Türkçe formunun geçerli ve güvenilir bir ölçek olarak ortaokul öğrencilerinin fen derslerindeki katılımını farklı boyutlarda ölçmek amacıyla kullanılabileceğini önermektedir. Öğrenci katılımı, başarı ile ilgili çıktılarla yüksek oranda ilişkili bulunduğu için bu ölçekten elde edilen veriler öğrencilerin fen başarılarını geliştirmek için tanılayıcı bir araç olarak kullanılabilir. Ancak ölçeğin sadece fen bilgisi alanında geçerli olduğunun unutulmaması gerekir çünkü öğrencilerin katılım seviyesi farklı alanlarda farklılık gösterebilmektedir. Dolayısıyla, öğrencilerin alana özgü katılım seviyelerinin incelenmesi gerekmektedir. Bu yüzden, ölçeğin dış geçerliliğin artırılması için ortaokul düzeyindeki farklı alanlarda, ölçeğin geçerliliğine kanıt sağlanması gerekmektedir. Ayrıca, bu ölçek farklı sınıf düzeylerinde de başarı ile ilgili çıktıları geliştirmek amacıyla öğrenci katılımını değerlendirmek için geçerli kılınabilir.

INTRODUCTION

In recent years, researchers, educators, and policymakers have been increasingly maintaining enthusiasm for understanding and collecting data on student engagement (Fredricks, Blumenfeld, & Paris, 2004). Possible reason of this interest is that engagement shapes to students' academic retention, resilience, and achievement (Skinner, Kindermann, & Furrer, 2008) as well as it is considered as a solution to students' high level of boredom, low achievement, and high dropping out rates (Fredricks, Blumenfeld, & Paris, 2004).

Despite the growing interest in student engagement, there is little consensus among researchers about how to conceptualize and define it due to being a wide construct (Farmer-Dougan, & McKinney, 2001). Student engagement was formerly described through behaviors that are observable like participation (Natriello, 1984 as cited in Fredricks et al., 2011). Then, emotional or affective aspects were also included into conceptualization of engagement (Connell, 1990; Finn, 1989). For example, most of the researchers adopted the following definition: engagement is the behavioral intensity and emotional quality of a student's active participation in a task (Connell, 1990; Connell & Wellborn, 1991; Wellborn, 1991). Miserandino (1996) also indicated that definition of student engagement covers students' behaviors such as participation, having persistence, being curious and calm. These definitions contain feelings of belonging, attachment, and enjoyment, and hence it can be said that these researchers adopt the two-dimensional engagement model which contains emotion (e.g., attention, belonging, enjoy, and positive emotions) and behavior (e.g., involvement, exertion, and positive conduct). Recently, some researchers (Archaumbault et al., 2009; Fredricks et al., 2004; Jimerson, Campos & Grief, 2003; Wigfield et al., 2008) have proposed a three-component model which contains emotional, behavioral, and cognitive aspects (i.e., strategic thinking, self-regulation, mastering the knowledge, and strategy use). Although there are many types or subcategories of engagement in the literature, the most frequently used components of student engagement in the reviewed literature have been identified as cognitive, behavioral, and emotional engagement in recent years (e.g. Dunleavy, 2008; Fredricks et al., 2004; Jimerson et al., 2003; National Research Council, 2004). More recently, a four-component model for engagement was proposed by Reeve and Tseng (2011). This four-component model includes behavioral, emotional, cognitive and agentic dimensions. That is, they integrate a new aspect called agentic engagement to the three-component model. They also supported the idea that three component models present incomplete understandings and thus a full understanding of how students engage themselves can be provided by adding this dimension. More specifically, according to this model, behavioral engagement can be described as students' effort, involvement, persistence and on-task attention during the learning activity (Furrer & Skinner, 2003; Skinner & Belmont, 1993). Emotional engagement can be defined as students' positive feelings like interest, enthusiasm, and enjoyment towards the topic, lesson, tasks and learning process rather than anxiety, anger or apathy (Skinner & Belmont, 1993; Skinner, Kindermann, & Furrer, 2009). Cognitive engagement refers to the degree to which students plan, organize, and reorganize learning duties in terms of academic ones by using strategic and sophisticated learning strategies (Greene & Miller, 1996; Meece, Blumenfeld, & Hoyle, 1988). Agentic engagement is more likely related to students' proactive and constructive contributions in learning environment by explaining what s/he prefers, asking questions, communicating with teacher such as making suggestion or requesting a demand and so on (Reeve & Tseng, 2011; Reeve, 2013). In other words, it refers to the degree to which a student tries to enrich the instruction which s/he gets rather than just passively receive it as a given (Reeve, 2012). If a student is agentically engaged, he or she would respond with something useful, such as suggesting proposal, asking questions, asking for an explanation and an example, declaring his or her thoughts and needs, advising an objective to be achieved, requesting learning opportunities and resources, looking for opportunities for increasing personal interest to the lesson (Reeve, 2012).

Although, as indicated by aforementioned literature, there are many terms to conceptualize engagement the only common trait among all of these conceptualizations is that engagement is multidimensional construct and this multidimensionality results in quite variability in the content

of items measuring these dimensions. Therefore, it is important to determine the components which are unique constructs for engagement, and the model which describes engagement best, and finally the most valid, reliable and suitable measure.

Based on the study of Fredricks, and McColskey (2012) in which the self-report measures of engagement were compared in terms of definition, usage, samples and psychometric information; it was thought that positive and negative sides of each measure can be determined and the most suitable instrument to assess student engagement can be identified for this study. With this aim, firstly, when the measures that have been used as measures of engagement are examined, it is seen that some scales are not direct measures of engagement but indirect measures. In other words, some instruments were saliently developed to assess engagement, whereas some were developed to assess the similar concepts such as motivation, identification with school, self-regulation, and strategy use. For instance, the Identification with School Questionnaire (ISQ) (Voelkl, 1996) measured sometimes student identification with school and sometimes emotional engagement from study to study. Likewise, the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich & DeGroot, 1990) was used both to self-regulation and strategy use and cognitive engagement. Similarly, "psychological engagement" subscale of the Student Engagement Instrument (SEI) (Appleton et al., 2006) contains items related to students' relationships with peers and teachers and support for learning from families, however, these items are indirect measures of engagement, not direct measures. Actually, engagement is not a construct that should be assessed indirectly. A measure which is designed for only assess the student engagement directly should be used to get more reliable results from the research especially on which the links between engagement and the other academic outcomes are investigated. Secondly, the elimination was made based on degree to which instrument represent the multidimensional nature of student engagement. When the instruments are examined based on this criterion, it is seen that there are only three instruments (i.e., High School Survey of Student Engagement (HSSSE) (Yazzie-Mintz, 2007), School Engagement Measure (SEM) (Fredericks, Blumenfeld, Friedel & Paris, 2005), and Engagement Questionnaire (EQ) (Reeve & Tseng, 2011) that includes explicitly behavioral, emotional and cognitive engagement dimensions. Concerning intended purpose of these instruments, SEM was designed to determine the link between classroom context and elementary students' engagement (Fredricks et al., 2005). HSSSE was developed to determine high school students' attitudes and perceptions towards their work, school learning environment, and their interaction with the school community (Yazzie-Mintz, 2007). EQ was developed with the aim of assessing four aspects of student engagement–behavioral, emotional, cognitive and agentic to find out a more fully understand how students engage themselves in learning activities. When the reliabilities of these instruments were considered, it was seen that the developers of SEM reported Cronbach's alpha of .72–.77 for behavioral engagement, .83–.86 for emotional engagement, and .55–.82 for cognitive engagement, whereas developers of EQ reported .94 for behavioral engagement, .78 for emotional engagement, .88 for cognitive engagement and .82 for agentic engagement. However, no reliability information related to HSSE was found. Concerning validity issue, Reeve and Tseng evaluated the fit of the 12 possible models for the EQ and they reported that among a single-factor model, all possible two-factor models, all possible three-factor models, and a four-factor, the best fit was shown by the four-factor model including behavioral, emotional, cognitive and agentic components. Moreover, Reeve and Tseng suggested that agency should be added as a new aspect of engagement, since it is the one thing that could explain the variance in students' achievement, and it completes the description of engagement and so fills the gap of the description of student engagement.

Additionally, there are also some studies which concluded that each of these components in this model is positively linked to students' achievement and self-efficacy. For example, behavioral engagement (i.e., attendance, effort, and insistence) was proposed to be positively related to self-efficacy (Bandura, 1997). Similarly, Linnenbrink and Pintrich (2003) reported that students who have confidence in their capabilities are inclined to make effort hard, continue to do, and look for help, otherwise they tended to give up easily when they meet with difficulties or get help to just complete the task, not to learn. In addition, higher levels of self-efficacy of students caused more positive emotions in academic contexts such as pride or happiness (Harter, 1992). On the other

hand, negative emotions such as anxiety and depression rise when students feel lack of self-efficacy. (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Harter, 1992; Meece, Wigfield, & Eccles, 1990; Pintrich & De Groot, 1990; Pintrich, Roeser, & De Groot, 1994). In terms of cognitive engagement, Pintrich and De Groot (1990) revealed that students with high self-efficacy tended to use diverse cognitive, metacognitive or self-regulatory learning strategies. In other study, Linnenbrink and Pintrich (2003) reported that if students have confidence in their skills, then they tend to strive to understand their tasks; to think deeply about it; to use deeper processing strategies over time such as elaboration, organizational strategies, and metacognitive strategies. Finally, Reeve and Lee (2014) revealed that high level of interest and self-efficacy can be considered as precursor of later obtaining of students' agentic engagement such as proactive contributions into the learning environment.

Consequently, the EQ developed based on the four-component model of student engagement (i.e., behavioral, emotional, cognitive and agentic engagement) appears to provide a valid and reliable measure of student engagement. Accordingly, the current study aimed to adopt the EQ to Turkish to assess student engagement.

At this point it is important to note that each of the dimensions of student engagement should be assessed specifically in one domain, because engagement and motivation issues are specific to domains such as Mathematics, English and Science and thus, global efforts to enrich academic engagement may not be as effective as efforts to increase domain specific engagement and motivation (Green, Martin & Marsh, 2007; Martin, 2008). Therefore, subject-specific measures are more powerfully linked to the corresponding subject-specific engagement and motivation constructs (Green, Martin & Marsh, 2007). Similarly, the study of the Martin's (2008) brings to the fore the significance of domain specificity for assessment. According to his findings, motivation and engagement are not domain general, rather they are domain specific. Therefore, assessment should be more target oriented. When this domain specificity is considered, this study focuses specifically on science domain and intends to validate the EQ to assess student engagement only in science domain. Future studies can attempt to validate the instrument for other domains such as math and language as well.

Assessment of student engagement is important because engagement is found to be related to students' achievement (Fredricks et al., 2004; Marks, 2000); critical thinking skills (Pascarella, Palmer, Moye, & Pierson, 2001), and problem solving skills (Johnson, Johnson, & Smith, 1998); to be an indicator of students' motivation during instruction (Furrer & Skinner, 2003; Skinner & Belmont, 1993). Therefore, teachers and researchers should be aware of this construct, determine the present condition and try to enhance student engagement.

METHOD

Participants

Two separate samples were used to validate the EQ for Turkish middle school students. Samples were selected using convenient sampling accompanied with cluster random sampling procedure: Districts from which the samples to be selected was chosen using convenient sampling. Then, within each district, public schools considered as clusters were selected randomly. Middle school students from these schools constituted each sample. Samples ranged in age from 12 to 15 years. Sample 1 which consisted of 153 students (68 boys and 85 girls) was used to obtain preliminary findings concerning factor structure and internal consistency of Turkish version of the Engagement Questionnaire conducting confirmatory factor analysis and reliability analyses. Sample 2 which consisted of 744 students (403 girls and 337 boys) was used to examine the psychometric properties of Turkish version of the Engagement Questionnaire in detail. For this purpose, confirmatory factor analysis and reliability analyses were again conducted. Pattern and structure coefficients were investigated and invariance across gender was explored. In addition, to provide further validity evidence for the questionnaire, the relation of each of its sub-scale with self-efficacy was examined.

Data Collection Tools

Engagement Questionnaire (EQ)

The EQ (See Appendix), developed by Reeve and Tseng (2011), was used to assess student engagement in terms of behavioral, emotional, cognitive and agentic engagement. It is a seven-point-Likert type self-report instrument, and it has 22 items in four sub-scales. During its development, items from different instruments were utilized, and it was applied to 369 high school students. To assess behavioral engagement, five-item measure was obtained from Miserandino's (1996) task involvement questionnaire, and this measure showed high reliability (e.g., "I listen carefully in class", α = .94). To assess emotional engagement, four items were selected from Wellborn's (1991) conceptualization of emotional engagement, and this measure showed sufficient reliability (e.g., "When I am in class, I feel curious about what we are learning", α = .78). To assess cognitive engagement, Wolters' (2004) learning strategies questionnaire was revised and new combined eight-item scale demonstrated high reliability (e.g., "When I study, I try to connect what I am learning with my own experiences", α = .88). Finally, to assess agentic engagement five-item measure was developed by Reeve and Tseng (2011). This measure showed sufficient reliability (e.g., "I offer suggestions about how to make the class better", α = .82). In addition, researchers conducted an exploratory factor analysis to test whether agentic engagement is a distinct aspect of engagement or not. Results confirm the four factors according to eigenvalue >1, and these four factors accounted for 66.6% of the total variance in the instrument. All factor loadings were .30 or greater. A series of confirmatory factor analyses were also conducted to determine the most appropriate structure for engagement by comparing all possible models of how the various aspects of engagement might combine to define a best-fitting structure. For this purpose, 12 possible models including a single-factor model in which all 22 indicators loaded onto a single latent variable; all possible two-factor models; all possible three-factor models; and four-factor model comprising four separate latent factors (i.e., behavioral, emotional, cognitive, and agentic) were tested. CFA results showed that all models fit the data based on to the chi-square statistic (all ps < .01), however, four-factor model adequately fit the data based on all fit indices (SRMR, RMSEA, CFI, and AIC) (Reeve & Tseng, 2011).

In order to validate the instrument for Turkish elementary school students, it was translated into Turkish by the first author. During its translation, "in science class" statement was added to the items in scale. For example, in the behavioral engagement sub-scale, there was an item "I listen carefully in class", and this item was edited as "I listen carefully in science classes". An instructor from science education department at a large public university whose area of research involves student motivation and self-regulation in science examined the translated instrument for content validity. The instructor also checked the quality of items in terms of clearness, sentence structure, and comprehensiveness. After this process, support from Academic Writing Center was received for the grammar structure of the translation. In the directions of these expert opinions, the instrument was revised. In order to determine if the instrument is comprehensible enough for students or not, it was read by five students. Some small changes were made on some words. In addition, different from the original version of the EQ which was on 7-point scale, 4-point-likert scale was decided to be used in the current study based on the students' comments: The students stated that it is easier for them to understand and respond to the 4-point format. Indeed, according to Bourke and Frampton (1992), younger individuals can respond to scales with fewer categories more easily. Accordingly, 4-point response format was utilized for adaptation of the EQ for Turkish middle school students.

Motivated Strategies for Learning Questionnaire (MSLQ)

It is a self-report instrument developed by Pintrich, Smith, Garcia, and Mckeachie (1993) to measure various aspects of student motivation, cognition, and behavior. The MSLQ consists of eleven sub-scales designed to be modular to be used by the researchers according to the scope of their studies (Pintrich, et al. 1993). It was translated and adapted into Turkish by Sungur (2004). Within the scope of the current study, self-efficacy for learning and performance sub-scale of the MSLQ was used to assess students' self-efficacy in science (e.g., "I believe I will receive an excellent

grade in science classes, *n* = 8) was used to provide further validity evidence for the Turkish version of the EQ and the coefficient alpha was found to be .90.

RESULTS

Pilot Study

Confirmatory Factor Analysis

As part of the pilot study, Turkish version of the questionnaire was administered to 153 public middle school students (Sample 1). Confirmatory factor analysis (CFA) was conducted using LISREL 8.80 for initial examination of the factor structure of the EQ. The result of CFA indicated a good model fit (CFI = .98, RMSEA = .05, SRMR = .05). Table 1 presents Lambda-Ksi estimates (pattern coefficients) for the subscales of EQ. As shown in the table, all Lambda-X estimates analogous to factor loadings in exploratory factor analysis were sufficiently high and statistically significant.

Table1. *Lambda-ksi Estimates (Pattern Coefficients) in the Pilot Study*

Subscale	Indicator	Pilot study LX estimates
Agentic Engagement	q1	.83
	q2	.63
	q3	.69
	q4	.64
	q5	.69
Behavioral Engagement	q6	.86
	q7	.77
	q8	.84
	q9	.73
	q10	.86
Cognitive Engagement	q15	.67
	q16	.67
	q17	.66
	q18	.75
	q19	.75
	q20	.62
	q21	.65
	q22	.54
Emotional Engagement	q11	.76
	q12	.74
	q13	.73
	q14	.72

Reliability Analysis

Concerning reliability analyses, Cronbach's alpha coefficient was computed for each sub-scale as a measure of internal consistency. The item-total correlations ranging from .53 to .85 suggested that all items contribute to the variability well and all of the items should be included in their corresponding sub-scales without a need for revision. Cronbach's alpha coefficients were found to be .82 for agentic engagement, .92 for behavioral engagement, .84 for emotional engagement, and .86 for cognitive engagement.

Main Study

Confirmatory Factor Analysis

After the pilot study, a second CFA was carried out with a new sample of 744 middle school students (Sample 2) in order to make a detailed exploration of the proposed factor structures. In addition, with this larger sample, measurement invariance across gender was examined. Consistent with the results of pilot study, second CFA results showed a good model to data fit with comparable fit indices (CFI = .99, RMSEA = .05, SRMR = .04). Graham, Guthrie, and Thompson (2003) suggested that when CFA is conducted, both pattern and structure coefficients should be reported and interpreted in the presence of factor correlation. The reason behind their suggestion was that when factors correlate with one another, observed variables correlate with all factors. Both pilot study and main study results demonstrated the presence of factor correlations. Thus, both Lambda-Ksi estimates (pattern coefficients) and structure coefficients were determined and displayed in Table 2.

Table 2. *Pattern and Structure Coefficients in the Main Study*

Indicator	Agentic	Behavioral	Cognitive	Emotional
q1	.67	.41	.50	.44
q2	.68	.41	.50	.44
q3	.72	.44	.53	.47
q4	.73	.45	.54	.48
q5	.67	.41	.50	.44
q6	.49	.81	.61	.66
q7	.43	.70	.53	.57
q8	.49	.80	.60	.66
q9	.44	.72	.54	.44
q10	.49	.80	.60	.66
q15	.50	.51	.68	.56
q16	.51	.52	.69	.57
q17	.50	.50	.67	.56
q18	.50	.51	.68	.56
q19	.49	.50	.66	.55
q20	.49	.50	.66	.55
q21	.50	.50	.67	.56
q22	.46	.47	.62	.51
q11	.51	.64	.65	.78
q12	.47	.60	.61	.73
q13	.51	.64	.65	.78
q14	.40	.53	.54	.65

Note: Non-italicized numbers are the pattern coefficient for each item with its designated factor.
Italicized numbers are the structure coefficient of each item with its nondesignated factors.

As shown in the table all items had high correlations with their corresponding factors (non-italicized numbers) and they were also correlated with their non-designated factors (italicized

numbers) but to a lower extent. All pattern coefficients were statistically significant. The range of pattern coefficients provided evidence for convergent validity.

In addition, all the confidence intervals (±2 standard errors) around correlations between any two factors (phi) which did not contain 1, provided an evidence for discriminant validity (Anderson & Gerbing, 1988). The maximum phi coefficient was found to be .83 with a confidence interval of .79 to .87.

Reliability Analysis

As part of reliability analyses, exploration of item-total correlations ranging from .55 to .76 revealed that all items contribute to the total variability well. In addition, Cronbach's alpha values were found to be .82 for agentic engagement, .88 for behavioral engagement, .86 for cognitive engagement and .83 for emotional engagement. These reliability values were close to those found in the original version of the EQ and Turkish version of the EQ in the pilot study.

Measurement Invariance across Gender

After conducting second CFA and reliability analyses, measurement invariance across gender was examined. Demonstration of and evidence for invariance across samples, increase the generalizability of the instrument. In order to determine the invariance, first, unconstrained baseline model was tested. Second, metric (weak) invariance in which factor loadings were set to be invariant across gender was specified. Third, the factor loadings and factor covariances invariant model was tested. Fourth, besides factor loadings, and factor covariances, factor variances were specified as invariant and tested. Finally, factor loadings, factor covariances, factor variances, and error terms were specified as invariant. In the current study, changes in CFI rather than chi-square difference ($\Delta\chi^2$) were explored to test measurement invariance, examined because chi-square test is sensitive to sample size (Cheung & Rensvold, 2002; Chen, 2007). Guided by Cheung and Rensvold (2002) and Chen (2007) cut-off values of $\Delta CFI \leq .01$ were used and results supported measurement invariance across gender (see Table 3). All ΔCFI values were less than 0.01.

More specifically, according to the results, the unconstrained baseline model provided a good fit ($\chi^2/df = 2.56$, CFI = .980). Pattern coefficients of all items were significant across gender. The difference between baseline model and factor loading invariant model was non-significant ($\Delta CFI = .001$) supporting factor invariance. Similarly, the difference between baseline model and the model specifying the factor loadings and covariances to be invariant across gender was not significant ($\Delta CFI = .001$). In addition, the same as the previous findings, there was a trivial difference between the baseline model and the model constraining factor loadings, covariances, and variances invariant across gender ($\Delta CFI = .001$). Finally, there was a non-significant difference between the baseline model and the model constraining factor loadings, covariances, and variances and individual item error terms invariant across gender ($\Delta CFI = .002$). Overall, with cut-off values of $\Delta CFI \leq .01$, results provided evidence for the statistical invariance of factor loadings, factor covariances, factor variances and individual item error terms.

Table 3. *Measurement Invariance of the EQ*

Model	χ^2/df	CFI
Baseline	2.56	.980
Factor loadings invariant	2.52	.979
Factor loadings and factor covariances invariant	2.52	.979
Factor loadings, factor covariances and factor variances invariant	2.52	.979
Factor loadings, factor covariances, factor variances, and error terms invariant	2.60	.977

Further Validity Evidence

Based on the available literature, positive correlations were expected between self-efficacy and engagement. Bivariate correlations revealed that self-efficacy was significantly and positively related to agentic ($r = .53$), behavioral ($r = .57$), cognitive ($r = .60$), and emotional engagement ($r = .58$).

DISCUSSION

The purpose of the present study was to validate the EQ to be used for Turkish middle school students in science classes. For this purpose, the EQ consisting of 22 items in four dimensions was first, translated into Turkish by the first author and considering expert opinions and middle school students' suggestions necessary changes were made in wording and scale format: The original questionnaire items was on a 7-point Likert scale. However, students commented that it is easier for them to respond to 4-point format. Actually, Bourke and Frampton (1992) reported that younger individuals can respond to scales with fewer categories more easily. In addition, a scale involving finer discrimination than the respondents can accomplish may cause a measurement error (Symonds, 1924 as cited in Weng, 2004). Accordingly, if the scales with many categories such as 7-point scales are used for young students who may not discriminate well among the categories, a measurement error can be introduced. On the other hand, too few categories can lead to less variability lowering reliability. For example, working with students with a mean age of approximately 12 years, Halpin, Halpin and Arbet (1994) found that changing the true-false format to 4-point Likert scale increased the estimates of Cronbach coefficient alpha. Thus, 4-point format appears to be more appropriate for middle school students. Accordingly, 4-point response format was decided to be used for adaptation of the EQ for Turkish middle school students.

Then, as part of the pilot study, Turkish version of the instrument was administered to 153 middle school students (Sample 1). CFA results supported the proposed factor structure. Moreover, reliability coefficients all above .80 demonstrated good internal consistency. Additionally, item-to-total correlations ranged from .53 to .85. Researchers in the field used different decision rules of item-to-total correlations for retention or deletion of items: For example, some retained items with item-to-total correlations greater than .35 (Bearden, Hardesty, & Rose, 2001), while others retained the items with item-to-total correlations greater than .50 (Obermiller & Spangenberg, 1998). When all these rules were considered, item-to-total correlations found in the current study which were all greater than .50 suggested that all items should be retained in their designated factors.

After pilot study, a second CFA was conducted with a larger sample (Sample 2, $n = 744$) to investigate the proposed factor structures in detail. Additionally, measurement invariance across gender was explored. The fit indices found in the second CFA were comparable with those found in the pilot study. Thus, the results concerning the validity of the proposed factor structure of the EQ based on the CFA results were consistent across pilot and main studies. In addition, all pattern coefficients obtained from both CFAs were high and statistically significant suggesting convergent validity. Examination of structure correlations, on the other hand, revealed that items have correlation with their non-designated factors as well. Some of the correlations were high which may suggest that discriminant validity was not well established. However, when the confidence intervals (±2 standard errors) around correlations between any two factors (phi) were examined as an evidence for discriminant validity, it was found that none of the intervals contained 1. This finding provided an evidence for discriminant validity.

Regarding reliability of sub-scales, similar with the findings from the pilot study, the reliability coefficients obtained in the main study, suggested high internal consistency and all items were found to contribute to the total variability well with item-to-total correlations ranging from .55 to .76.

The results concerning the measurement invariance across gender revealed evidence for the statistical invariance of factor loadings, factor covariances, factor variances and individual

item error terms. This finding implies that the items are likely to function in a similar way for boys and girls contributing to the generalizability of the instrument. Accordingly, it is expected that the data from both genders lead to equally valid conclusions regarding their engagement levels.

In order to provide further validity evidence for the EQ, bivariate correlations between self-efficacy and each of the EQ sub-scale were determined. As expected, positive and significant correlation were found. Indeed, relevant literature demonstrated that self-efficacy has an important role in student engagement (Linnenbrink & Pintrich, 2003). Actually, self-efficacious students are likely to persist in the face of challenging situations (Bandura, 1997), use various strategies to achieve their goals (Walker, Greene, & Mansell, 2006), have positive emotions (Harter, 1992), and make proactive contributions to the learning environment (Reeve & Lee, 2014).

Overall, the present study suggests that Turkish version of the EQ can be used as a valid and reliable measure to assess different aspects of middle school students' engagement in science. Because, engagement is found to be highly correlated with achievement related outcomes (Appleton, Christenson, & Furlong, 2008; Avenilla, 2003; Miller, Greene, Montalvo, Ravindran, & Nichols, 1996), the data from this instrument can be used as diagnostic tool to improve students' achievement in science.

At this point, it is important to note that the EQ was validated only for science domain because according to relevant literature students' engagement level may show differences in different domains (Green, Martin & Marsh, 2007; Martin, 2008). Accordingly, it is essential to examine students' domain specific engagement levels. So, in order to increase external validity of the EQ, there is a need for providing validity evidences for the instrument across different domains at middle school level. In addition, the instrument can be validated for students at different schooling levels to assess their engagement with an ultimate aim of improving achievement related outcomes.

REFERENCES

Anderson, J.C. & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411-423.

Appleton, J. J., Christenson, S. L., & Furlong, M. 1. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools, 45*(5), 369-386.

Archambault, I., Janosz, M., Fallu, J. S., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of adolescence*, *32*(3), 651-670.

Archambault, I., Janosz, M., Morizot, J., & Pagani, L. (2009). Adolescent behavioral, affective, and cognitive engagement in school: Relationship to dropout. *Journal of school Health*, *79*(9), 408-415.

Avenilla, F. R. (2003). *Assessing the links between emotional and behavioral school engagement and academic outcomes among high school students.* Unpublished doctoral dissertation, The Pennsylvania State University.

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

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development*, *67*(3), 1206-1222.

Bearden, W. O., Hardesty, D.M, & Rose, R.L. (2001). Consumer self-confidence: Refinements in conceptualization and measurement. *Journal of Consumer Research, 28*, 121–34.

Bourke, S. & Frampton, J. (1992, November). *Assessing the quality of school life: Some technical considerations.* Paper presented at the Annual Conference of the AARE, Deakin University, Geelong, Australia..

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

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

Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2012). *Handbook of research on student engagement.* Springer Science & Business Media.

Connell, J. P. (1990). Context, self, and action: A motivational analysis of self-system processes across the life span. *The self in transition: Infancy to childhood*, 61-97.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar & L. A. Sroufe (Eds.), *Self-processes and development* (pp. 43-77). Hillsdale, NJ: Lawrence Erlbaum.

Dogan, U. (2014). Validity and Reliability of Student Engagement Scale. *Bartin Üniversitesi Egitim Fakültesi Dergisi*, *3*(2), 390.

Dunleavy, J., & Milton, P. (2008). Student Engagement for Effective Teaching and Deep Learning. *Education Canada*, *48*(5), 4-8.

Farmer-Dougan, V., & McKinney, K. (2001). Examining student engagement at Illinois State University: An exploratory investigation. *Center for Teaching, Learning & Technology.* Illinois State University

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*(2), 117-142.

Fredericks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*, 59-109.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of research on student engagement* (pp. 763-782). Springer US.

Fredricks, J. A., Blumenfeld, P. C., Friedel, J., & Paris, A. (2005). School engagement. In K. A. Moore & L. Lippman (Eds.), *conceptualizing and measuring indicators of positive development: What do children need to flourish* (pp. 305–321). New York: Kluwer Academic/Plenum Press

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments. Issues & Answers. REL 2011-No. 098. *Regional Educational Laboratory Southeast.*

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of educational psychology*,*95*(1), 148.

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007). Between-and within-domain relations of students' academic emotions. *Journal of Educational Psychology*, *99*(4), 715.

Graham, J. M., Guthrie, A. C., & Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: A reminder. *Structural Equation Modeling, 10,* 142-153.

Green, J., Martin, A. J., & Marsh, H. W. (2007). Motivation and engagement in English, mathematics and science high school subjects: Towards an understanding of multidimensional domain specificity. *Learning and Individual Differences*, *17*(3), 269-279.

Greene, B. A., & Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology*, *21*(2), 181-192.

Halpin, G., Halpin, G., & Arbet, S. (1994). Effects of number and type of response choices on internal consistency reliability. *Perceptual and Motor Skills*, *79*(2), 928-930.

Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, *43*(9), 602-615.

Jimerson, S. R., Campos, E., & Greif, J. L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist*, *8*(1), 7-27.

Johnson, D. W., Johnson, R., & Smith, K. (1998).Active learning: Cooperation in the college classroom. Edina, MN: Interaction Book Company.

Kahraman, N. (2014). Cross-Grade Comparison of Relationship between Students' Engagement and TIMSS 2011 Science Achievement. *Egitim ve Bilim*, *39*(172).

Ladd, G. W., & Dinella, L. M. (2009). Continuity and change in early school engagement: Predictive of children's achievement trajectories from first to eighth grade?. *Journal of Educational Psychology*, *101*(1), 190.

Linnenbrink, E. A. & Pintrich, P.R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly, 19,* 119-137.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal*, *37*(1), 153-184.

Martin, A. J. (2008). Enhancing student motivation and engagement: The effects of a multidimensional intervention. *Contemporary Educational Psychology*, *33*(2), 239-269.

Mazer, J. P. (2013). Associations among teacher communication behaviors, student interest, and engagement: A validity test. *Communication Education*,*62*(1), 86-96.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology*, *82*(1), 60.

Meece, J., Blumenfeld, P. C., & Hoyle, R. H. (1988).Students' goal orientation and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*, 514–523.

Miller, R. B., Greene, B. A., Montalvo, G. P., Ravindran, B., & Nichols, J. D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology*, *21*(4), 388-422.

Miserandino, M. (1996). Children who do well in school: Individual differences in perceived competence and autonomy in above-average children. *Journal of Educational Psychology, 88,* 203-214.

National Research Council and the Institute of Medicine. (2004). *Engaging schools: Fostering high school students' motivation to learn*. Committee on Increasing High School Students' Engagement and Motivation to Learn. Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Natriello, G. (1984). Problems in the evaluation of students and student disengagement for secondary schools. Journal of Research and Development in Education, 17, 14–24.

Obermiller, C., & Spangenberg, E.R. (1998). Development of a scale to measure consumer skepticism toward advertising. *Journal of Consumer Psychology, 7,* 159 – 186.

Pascarella, E. T., Palmer, B., Moye, M., & Pierson, C. T. (2001). Do diversity experiences influence the development of critical thinking? *Journal of College Student Development*, 42, 257-271.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33.

Pintrich, P. R., Roeser, R. W., & De Groot, E. A. (1994). Classroom and individual differences in early adolescents' motivation and self-regulated learning. *The Journal of Early Adolescence*, *14*(2), 139-161.

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801-813.

Reeve, J. (2012). A Self-determination Theory Perspective on student Engagement. In S.L. Christenson et al. (eds.), *Handbook of research on student engagement*. (pp. 149-171). New York, NY: Springer.

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology, 105,* 579-595. doi:10.1037/a0032690

Reeve, J., & Lee, W. (2014). Students' classroom engagement produces longitudinal changes in classroom motivation. *Journal of Educational Psychology*, 106, 527-540.

Reeve, J., & Tseng, C.-M (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, *36*, 257-267.

Reschly, A. L., & Christenson, S. L. (2006). Prediction of dropout among students with mild disabilities: A case for inclusion of student engagement variables. *Remedial and Special Education, 27,* 276–292.

Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2008). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*.

Skinner, E. A., Kindermann, T. A., Connell, J. P., & Wellborn, J. G. (2009). Engagement and disaffection as organizational constructs in the dynamics of motivational development. *Handbook of motivation at school*, 223-245.

Skinner, E., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effect of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85,* 571–581.

Sungur, S. (2004). *An implementation of problem based learning in high school biology courses* (Doctoral dissertation, Middle East Technical University).

Uğur, E., & Akın, A. (2015). The Psychometric Properties of Turkish Version of the Student Engagement Scale. *SDU International Journal of Educational Studies*, *2*(1), 53-59.

Voelkl, K. E. (1996). Measuring students' identification with school. *Educational and Psychological*

Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences*, *16*(1), 1-12.

Wellborn, J. G. (1991). *Engaged and disaffected action: The conceptualization and measurement of motivation in the academic domain.* (Unpublished doctoral dissertation). University of Rochester, Rochester, NY.

Weng, L.J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement, 64,* 956-972.

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, *45*(5), 432-445.

Wolters, C. A. (2004). Advancing Achievement Goal Theory: Using Goal Structures and Goal Orientations to Predict Students' Motivation, Cognition, and Achievement. *Journal of educational psychology*, *96*(2), 236.

Yazzie-Mintz, E. (2007). *Voices of students on engagement: A report on the 2006 High School Survey of Student Engagement.* Bloomington, IN: Center for Evaluation & Educational Policy, Indiana University.

Appendix- Engagement Questionnaire

	Kesinlikle katılmıyorum	Katılmıyorm	Katılıyorum	Kesinlikle Katılıyorum
1) Fen dersinde öğretmenime sorular sorarım.	1	2	3	4
2) Dersle ilgili sevdiğim ya da sevmediğim şeyleri fen öğretmenime söylerim.	1	2	3	4
3) Fen dersiyle ilgili nelere ilgi duyduğumu öğretmenime söylerim.	1	2	3	4
4) Fen dersiyle ilgili tercihlerimi ve düşüncelerimi açıkça ifade ederim.	1	2	3	4
5) Fen dersini daha iyi hale getirebilmek için önerilerde bulunurum.	1	2	3	4
6) Fen dersini dikkatle dinlerim.	1	2	3	4
7) Fen dersine çok çalışırım.	1	2	3	4
8) Fen öğretmenimiz yeni bir konuya başladığında, dikkatle dinlerim.	1	2	3	4
9) Fen dersinde yeni bir konuya başladığımızda, çok çalışırım.	1	2	3	4
10) Fen dersine dikkatimi veririm.	1	2	3	4
11) Fen dersinde yeni şeyler öğrenmekten hoşlanırım.	1	2	3	4
12) Fen dersinde herhangi bir şey üzerinde çalışmak ilgimi çeker.	1	2	3	4
13) Fen dersinde öğrendiklerimize karşı merak duyuyorum.	1	2	3	4
14) Fen dersi eğlencelidir.	1	2	3	4
15) Fen dersindeki yeni bilgileri eski bilgilerimle ilişkilendirmeye çalışırım.	1	2	3	4
16) Fen dersine çalışırken yeni bilgilerle kendi deneyimlerim arasında bağlantı kurmaya çalışırım.	1	2	3	4
17) Fen dersine çalışırken tüm farklı fikirleri bir araya getirerek, onları anlamlandırmaya çalışırım.	1	2	3	4
18) Fen dersine çalışırken, kendi örneklerimi oluşturarak önemli kavramları anlamaya çalışırım.	1	2	3	4
19) Fen dersine çalışmaya başlamadan önce, ulaşmak istediğim hedefi belirlerim.	1	2	3	4
20) Fen dersine çalışırken, ara sıra durur, yaptıklarımı gözden geçiririm.	1	2	3	4
21) Fen dersine çalışırken, yalnızca doğru cevapları bulup bulamadığıma değil, ne kadar anladığıma da dikkat ederim.	1	2	3	4
22) Eğer bir fen konusunu anlamakta zorlanıyorsam, onu öğrenmek için izlediğim yolu değiştiririm.	1	2	3	4