ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
2020; 11(2); 163-179

Adaptation of the Self-efficacy Beliefs in STEM Education Scale and Testing Measurement Invariance across Groups *

Cansu DEMİRBAĞ ** Serkan ARIKAN *** Ebru Zeynep MUĞALOĞLU ****

Abstract
Academic performance on science, technology, engineering, and mathematics (STEM) education is important for the economic development of countries. From the perspectives of social cognitive theory, one of the predictors of academic performance is self-efficacy. In order to measure middle school students' self-efficacy beliefs in STEM education, STEM Competency Beliefs scale was developed in English originally by Chen, Cannady, Schunn, and Dorph (2017). In this study, it is aimed to adapt the English scale into Turkish and to provide evidence regarding reliability and validity. Throughout the adaptation process, forward and backward translation was completed. In the pilot study (n = 77), the reliability of the data and the clarity of the statements in the Turkish version of the scale was examined. In the main study, the Turkish version was administered to 330 middle school students to investigate the psychometric properties of the scale. The results pointed out that the scores obtained by the Turkish version of the scale had good internal consistency. Regarding the dimensionality of the scale, in contrast to the original version, the adapted scale showed a two-dimensional structure. Measurement invariance findings for gender groups supported configural and metric invariance, whereas scalar invariance was partially achieved. Measurement invariance findings for career choice groups supported configural, metric, and scalar invariance. Scale scores of students were estimated using multidimensional Item Response Theory. The findings suggested that the scale can be utilized for STEM-related research to assess the competency beliefs of students.

Key Words: Self-efficacy beliefs, scale adaptation, confirmatory factor analysis, measurement invariance, multidimensional item response theory.

INTRODUCTION

Science, technology, engineering, and mathematics (STEM) education is the integration of these disciplines (Breiner, Harkness, Johnson, & Koehler, 2012; Tsupros, Kohler, & Hallinen, 2009) in order to deal with real-world problems (Johnson, Peters-Burton, & Moore, 2016; National Research Council-NRC, 2014). STEM education is substantial for countries in terms of three interconnected aspects: competitiveness in the global market, needs for innovation, and jobs of the future (Atkinson & Mayo, 2010; English, 2016; Johnson et al., 2016). One of the ways to stay competitive in global markets for countries is maintaining development in STEM disciplines. Science- and technology-based innovation enforces countries in the global market by increasing exports (Atkinson & Mayo, 2010). This kind of innovation is only possible with a workforce educated in science, technology, engineering, and mathematics content (Atkinson & Mayo, 2010). It is predicted that in the future one out of three jobs will be STEM-integrated or strongly related to STEM fields. Hence, students need to be educated with integrated STEM approach as candidates for the future workforce (English, 2016).

Similarly, Turkey, as a developing country, emphasizes the importance of STEM education for its' economic growth (TÜSİAD, 2019). Turkey needs a qualified and talented workforce educated through

* This study is a part of Master Thesis entitled "A Turkish Adaptation of The Stem Competency Beliefs Instrument" completed within Boğaziçi University Institute for Graduate Studies in Social Sciences.

** Graduate Student, Boğaziçi University, Faculty of Education, İstanbul-Turkey, cansu.demirtas@boun.edu.tr, ORCID ID: 0000-0001-9072-0375
*** Assist. Prof., Boğaziçi University, Faculty of Education, İstanbul-Turkey, serkan.arikan1@boun.edu.tr, ORCID ID: 0000-0001-9610-5496
**** Assoc. Prof., Boğaziçi University, Faculty of Education, İstanbul-Turkey, akturkeb@boun.edu.tr, ORCID ID: 0000-0001-7766-4743

To cite this article:
Demirbağ, C., Arıkan, S., & Muğaloğlu, E. Z. (2020). Adaptation of the self-efficacy beliefs in STEM education scale and testing measurement invariance across groups. *Journal of Measurement and Evaluation in Education and Psychology, 11*(2), 163-179. doi: 10.21031/epod.675240

Received: 15.01.2020
Accepted: 26.04.2020

STEM fields to achieve the goals of 2023. Preliminary actions have been done, such as changing the national curriculum (Ministry of National Education-MEB, 2018a) and opening STEM institutions and centers to empower STEM education (Colakoğlu & Gökben, 2017). Moreover, research about STEM studies and developing STEM-related master and doctorate programs have been increasing (Akgündüz et al., 2015).

Self-efficacy beliefs are regarded as one of the variables that play a key role in academic achievement (Jinks & Lorsbach, 2003; Kanny, Sax, & Riggers-Piehl, 2014; Nelson & Ketelhut, 2008) and career persistence (Green & Sanderson, 2018) in STEM fields. It is significant to improve self-efficacy and academic achievement of students in STEM fields to fulfill the STEM-related jobs. Even though the number of STEM education research has gained acceleration both at international level (Atkinson & Mayo, 2010; Breiner et al., 2012; English, 2016; Johnson et al., 2016; Tsupros et al., 2009) and in Turkey (Han, Capraro, & Capraro, 2016; Hacıoğlu, Yamak, & Kavak, 2016; Yerdelen, Kahraman, & Taş, 2016), to the best of our knowledge, there is not a valid scale to assess the STEM self-efficacy beliefs in Turkey.

Firstly, the present study aimed at adapting the English version of the STEM Competency Beliefs scale into Turkish and validating the adapted version. Secondly, the study compared the participants' self-efficacy beliefs on STEM education in terms of their gender, school type, and career choices in a Turkish context. Finding significant differences between school types (private vs. public) and career choices (stem related and not-stem related) could be considered as additional validity evidence (Sireci & Sukin, 2013) as these groups are expected to be different in their competency scores due to the resources and student motivation, respectively.

Having a valid scale to assess STEM self-efficacy beliefs in Turkish is significant for researchers and educators to investigate individual's self-efficacy on STEM and its relationships with other crucial variables such as academic performance in STEM or interest towards STEM fields in Turkey. Moreover, having a STEM Competency Belief scale in Turkish enables researchers, teachers and policymakers to evaluate STEM programs and identify the learner characteristics in terms of STEM self-efficacy in Turkey. Comparing STEM competency beliefs of gender groups in Turkey is also expected to extend the literature.

Self-efficacy Beliefs in STEM Education

Self-efficacy is defined as the capability of an individual's point of view for himself/herself to perform at a level of proficiency (Bandura, 1999) and interchangeably used perceived self-competence (Zimmerman, 1995). Self-efficient people are more resilient, solution-oriented, hard workers (Pajares & Miller, 1997), active in the control of time, better at task focus (Bouffard-Bouchard, Parent, & Larivee, 1991), self-regulated, more efficient in the use of problem-solving strategies and in the management of working time (Zimmerman, 2000). Bandura (1999) also explained that self-efficient people perceived failure differently than less self-efficient people. They regard failure to insufficient effort, weak strategies, or conditions. These features of self-efficient people play a key role in their performance (Bandura, 1999; Bouffard-Bouchard et al., 1991).

Beliefs about self-efficacy influence how much students learn (Vincent-Ruz & Schunn, 2017). For instance, Nelson and Ketelhut (2008) investigated ninety-six middle school students' self-efficacy and their performance in learning science in a virtual environment. As a result of the study, it was indicated that students with lower levels of self-efficacy did not perform as well as students with higher levels of self-efficacy. Bandura (1997) emphasized that the relationship between self-efficacy and performance is reciprocal. In other words, if people are self-efficient, their characteristics help them to be successful in related tasks. Achieving tasks boosts their self-efficacy, which leads to working harder and targeting more difficult tasks. Working harder helps to achieve new tasks that continue with better performance and higher self-efficacy. Moreover, Hidi and Ainley (2008) emphasized a positive relationship between interest and self-efficacy. The more students believe themselves, the more they are interested in their subjects. Thus, educators are required to help learners to experience better feelings and improve their beliefs about themselves. It helps students continue to work on or reengage

with activities, ideas, objects and so on, and to increase knowledge and a stored value (Hidi & Ainley, 2008).

Beliefs about capabilities function as an important role that influences science or non-science related majors and career choices (Hackett & Betz, 1982). Durik, Vida, and Eccles (2006) examined how the 10th graders' self-concept of ability on English/reading was related to their career choices. The results showed that the subject-oriented self-concept of ability predicted future career preferences of 10th graders. Gainor (2006) also emphasized that people choose careers in areas where they believe that they are good at doing it well.

Studies found that females have lower self-efficacy towards STEM fields (Tellhed, Backström, & Björklund, 2017). Females do not believe that they can accomplish STEM fields because of the lack of role models and social or verbal persuasions (Zeldin, Britner, & Pajares, 2008). Self-doubts, lower performance expectations, male-dominated fields, social persuasions and vicarious experiences about STEM fields, individual backgrounds, family influences and expectations, perceptions towards STEM fields, psychological values, factors, and preferences are related with females' lower interests towards STEM fields (Kanny et al., 2014; Tellhed et al., 2017; Zeldin et al., 2008). Lower self-efficacy beliefs of females towards STEM is needed to overcome to reduce gender segregation in the field. One of the ways for increasing females in the area is increasing their self-efficacy for STEM careers (Tellhed et al., 2017).

Self-efficacy is a personal state which can change especially based on positive personal outcomes. As Jenson, Petri, Day, Truman, and Duffy (2011) stated STEM self-efficacy is an important focus and worthy of observation. Therefore, to assess STEM self-efficacy, many scales have been developed over the years (e.g., Dawes, Horan, & Hackett, 2000; Lent, Brown, & Larkin, 1986). In 2014, Milner, Horan, and Tracey (2014) argued that most of the scales have validity issues, and they developed the STEM Career Self-Efficacy Test. Pieces of evidence were presented to claim that the scale can be accepted as a valid instrument to measure self-efficacy in engaging STEM activities (Milner et. al., 2014). However, the scale is not applicable to middle school students who are expected to learn STEM fields at schools. In 2017, the STEM Competency Beliefs scale was developed for middle school students in Activation Lab in the USA (Chen, Cannady, Schunn, & Dorph 2017). Activation Lab gathers academicians from various universities of the USA. They aim to increase young people's understanding and appreciation of STEM to prepare them for future challenges. One of the main research areas in Activation Lab is to develop scales to measure significant variables for STEM education, such as Science Competency Scale (Chung, Cannady, Schunn, Dorph, & Vincent-Ruz, (2016) and STEM Competency Belief scale (Chen et al., 2017). The STEM Competency Belief scale was developed to assess an individual's STEM Competency Beliefs. Cannady stated that the scale was also adapted into different languages like Spanish and African (M. Cannady, personal communication, November 12, 2018). As the original scale was developed very recently, there is not any publication yet based on this scale. Moreover, Smith (2019) adapted the original scale to measure technology competency beliefs. She applied the adapted version to investigate the effect of a coding instruction to seventh graders' self-efficacy in technology.

Present Study

In a decade when STEM has gained popularity and been studied from different perspectives, it is crucial to assess the self-efficacy of students for STEM fields. One of the scales to assess middle school students' self-efficacy in STEM education is the STEM Competency Beliefs scale. The scale was developed by Chen et al. (2017) in English. The purpose of the present study was twofold. First, to adapt the scale into Turkish and to test the factor structure of the STEM Competency Beliefs scale with the Turkish sample. The second purpose was to test whether the factor structure of the scale had measurement invariance across gender groups and career choice groups in the Turkish sample. The research questions of this study are:

1) Does the factor structure of the adapted STEM Competency Beliefs scale similar to the original scale?
2) Are the configural, metric, and scalar parameters invariant across girls and boys?
3) Are the configural, metric, and scalar parameters invariant across students who want to follow stem-related and not stem-related careers?
4) Is there any significant difference between students' scale scores on gender groups, career groups, and school types?

METHOD

This study primarily aimed to adapt STEM Competency Beliefs scale into Turkish and to test measurement invariance for the factor structure of the STEM Competency Beliefs scale. Therefore, the adaptation part could be named as a descriptive study and measurement invariance part could be named as a correlational study. Detailed information about participants, data collection instrument and data analysis are presented below.

Participants

For the pilot and the main study, two different sample groups were used. All the students were science center visitors taken by their schools as a school trip to attend workshops; therefore, the sampling method was the convenience sampling. These workshops were held in a science center in İstanbul which belongs to a Municipality. Seventy-seven students (4^{th} to 8^{th} graders) participated in the pilot study. The participants consisted of 32 male (42%) and 45 female (58%) students. Seven of the participants (9%) were from private schools, and 70 of them (91%) were from public schools.

Participants of the main study were 330 students coming from different schools as visitors to the science center. Among these 330 students, 4 of them did not provide all responses to the items. Therefore, after listwise deletion, all the analyses were conducted based on 326 students (2 females and 2 males; 3 public and 1 private school). The gender percentages of the students were regarded as balanced, consisting of 157 females (48%) and 169 males (52%). Also, students who participated in the study were coming from different school types as public schools (n = 302, 93%) and private schools (n = 24, 7%). The majority of the students were 7th graders. Among these students, 161 of them (49%) stated that they want to have STEM-related careers, whereas 165 of them (51%) do not want to follow STEM-related careers. According to student ratios of gender groups, school types, and students' choices of future careers, and the way these students were brought to the center, the sample could be considered as not biased.

Data Collection Instrument

The STEM Competency Belief scale is a 12-item 4-point Likert-type scale (Chen et al., 2017). The survey was designed for 10-14-year-old respondents to assess an individual's STEM Competency Beliefs. The reliability of the STEM Competency Beliefs Scale was good (Cronbach's Alpha = .83; polychoric Alpha = .87) based on a data collected from a sample of 205 middle school youth (Chen et al., 2017). Two of the items were listed below as sample items:

"I can do math problems I get in the class."

"I am the technology expert in the house."

Data Analysis

The scale adaptation process included the following stages: scale adaptation, piloting, reliability and validity analysis, and testing measurement invariance for gender groups and career choice groups.

Scale adaptation

Methodology in translation and adaptation of a scale has enhanced rapidly in last 25 years. The reasons behind this rapid development are based on four issues including interest in cross-cultural psychology (van de Vijver & Hambleton, 1996), international comparative studies in education, worldwide exams, and fairness in testing for language preferences (Hambleton, Merenda & Spielberger, 2012; International Test Commission-ITC, 2017;).

Translation and adaptation are two major terms used in the field. Compared to the test translation, the test adaptation is a more preferred, more reflective, broader, and commonly used term (Hambleton et al., 2012; ITC, 2017). During the application of test adaptation, a variety of activities are required, such as deciding whether the same construct occurs in different languages, determining translators, deciding accommodations, adapting the tests, and checking for equivalence. On the other hand, the test translation is only one of the steps that happen in the adaptation. This step is language translation from one to another. However, a test adaptation requires thinking deeply in terms of cultural, psychological and linguistic issues (Hambleton et al., 2012). Briefly, translation and adaptation have different meanings, and the adaptation is a more comprehensive term.

ITC (2017) guideline grouped the steps of the test adaptation process as before, in progress, and after. According to the guideline, before the adaptation, three steps are suggested for experts: obtaining permission from test developers, evaluating the similarities between cultures, and minimizing the cultural and linguistic differences. In the progress part of the adaptation, five steps are emphasized: ensuring the minimal cultural differences, using appropriate design methods to maximize suitability, providing evidence that the test is the same for intended populations, providing evidence for the structure of the test, collecting data to complete necessary revisions. In the last part, four steps are needed to be completed after the adaptation process: determining the sufficient size of the sample, providing statistical evidence for construct equivalence, providing evidence for reliability and validity analysis, and using appropriate data analysis procedure. In addition to the steps mentioned here, scoring and documentation are emphasized in the guideline (ITC, 2017).

For the adaptation process, two main design methods appear in the literature, namely forward and backward translation. The forward translation is a process that one or more translators adapt the test from the source language to the target language. Backward translation has three main processes in itself. Firstly, a test is translated from the source language to target language by determined translators. Then, different translators translate the test from target language back to the source language. Finally, these two forms of the test as source language and back-translated version are compared for equivalence (Hambleton et al., 2012). The backward translation allows the researcher to compare two forms in a more objective level.

For the adaptation of the STEM Competency Beliefs scale, preconditions were completed before the study. Firstly, permission was granted for the adaptation of the STEM Competency Beliefs scale into the Turkish (M. Cannady, personal communication, November 12, 2018). Then, cultural similarities and differences were evaluated by the research team, including an associate professor in science education, an assistant professor in assessment and evaluation, and the researcher. Finally, forward translation, backward translation, and final version editing were performed.

Forward translation: For the forward translation, the scale was translated from English to Turkish. Translators were 5 years experienced English teacher and 7 years experienced English interpreter. Each translator worked independently, and translated forms were collected in an excel document. The research team compared the translations, discussed STEM-related terms, and the scale was formed in Turkish. For example, the research team discussed "After school science club" and decided to translate as "science and technology club" which is a term in the National Education Social Activities Program Students' Club (MEB, 2009).

Backward translation: To achieve backward translation, two additional translators translated the scale from Turkish to English. These translators were a Turkish scholar who lived in England for 25 years and an American author who has been living in Istanbul for 14 years. Back-translated forms were

again collected in an excel document, and the research team investigated the similarities between the original form of the scale with back-translated form. After all, the research team reached a consensus for the back-translated scale.

Final version editing: As a final step, a linguist expert who is a doctorate student in a Learning Science program and a Turkish language editor compared the back-translated version of the scale and the original one. After some smooth changes on the adapted scale, the adapted Turkish version was finalized.

Piloting the adapted version of the scale

A pilot study was conducted to check the clarity of the items from students' perspectives. There were 2 additional questions at the end of the survey: "Is there any question that you struggle to understand?" and "if yes, which question(s) were they?" to identify problematic statements. Additionally, Cronbach's Alpha value and corrected item-total correlations were estimated to flag problematic items. Related revisions were made as a result of the pilot analysis.

Reliability analysis of final data

The reliability of the scale was tested using Cronbach's Alpha internal consistency coefficient. Cronbach's Alpha value above .70 is acceptable, above .80 is good, and .90 and above is excellent. Results that are closer to 1 mean higher internal consistency (George & Mallery, 2001). In the item level, the corrected-item total correlations were reported. Items with low correlations (less than .30) are considered as problematic items (Field, 2013), and these items are investigated to detect the source of the problem.

Validity analysis of the final data

For the validity analysis, confirmatory factor analysis (CFA) was conducted. CFA is one of the forms of factor analysis to test whether the hypothesized structure fits the collected data well or not (Urdan, 2010). In order to evaluate the goodness of the fit of the data for the proposed model, fit indices are used. CFI (Comparative Fit Index), TLI (Tucker Lewis index) and RMSEA (Root Mean Square Error of Approximation) are widely used fit indices that are less sensitive to the sample size. CFI and TLI values over .95 and RMSEA value smaller than .06 is accepted as a good fit (Ullman, 2001). CFA analysis for the study was conducted with MPLUS 7.4 (Muthén & Muthén, 2015) using the Weighted Least Square estimation method. One dimensional structure proposed in the English version was tested with the data collected by the adapted Turkish version. Multivariate normality, outliers, and sample size assumptions were checked to conduct CFA (Ullman, 2001).

When the student data does not fit the hypothesized structure, exploratory factor analysis (EFA) could be used to investigate the communalities among items. EFA using principal axis factor extraction technique with direct oblimin rotation was conducted as items could be correlated with each other. An item that has 0.400 or less item loading to its primary factor is considered as a problematic item. Also, if an item is loaded to at least two factors at the same time (factor loading difference of an item to a primary factor and other factor is less than .10), that item is also called problematic item (Field, 2013).

Item response theory scaling

Item response theory (IRT) scaling was conducted to estimate students' ability on the latent variables. Generally, IRT requires the data to be unidimensional (Hambleton & Jones, 1993). In the case of violating unidimensionality, multidimensional IRT estimations are available (Reckase, 2009). IRTPRO 4.2 (Cai, Thissen & du Toit, 2017) software was used to estimate the student ability as the

software is capable of conducting unidimensional and multidimensional IRT. Bock-Aitkin Expectation-Maximization estimation method was used.

Measurement invariance of final data

Measurement invariance analysis for gender groups and career choice groups were conducted to test whether the same construct was being measured across groups. As the number of students from private schools was not enough to estimate the parameters, measurement invariance analysis for school type was not performed. Having measurement invariance across gender or career choice groups implies that the scale scores of boys and girls, or students who want stem-related and not stem-related careers are comparable. The measurement invariance is tested comparing fit results of nested models: configural, metric, and scalar models. In the configural model, whether the same factor structure exists across groups is tested. In this model, factor loadings and thresholds are freed to be different across groups. In the metric model, factor loadings were constrained to be equal across groups, but the thresholds could take different values. In the scalar model, both factor loadings and item thresholds are constrained to be equal for groups (Milfont & Fischer, 2010; Vandenberg & Lance, 2000). Measurement invariance is assessed by comparing ΔCFI and ΔRMSEA values with cutoff criteria ($\Delta CFI \leq .01$, $\Delta RMSEA \leq .015$) suggested by Chen (2007), and Cheung and Rensvold (2002).

RESULTS

Pilot Study of the Scale

In the pilot study, items were administered to 77 students to test the clarity and fluency of the statements mainly. There were 2 additional questions at the end of the survey: "Is there any question that you struggle to understand?" and "if yes, which question(s) were they?" Seventy-two students stated that they could understand the statements clearly, and five students indicated that they had a problem to understand some items. These answers were used to determine if the statements need any changes or improvements before finalizing the Turkish version. For instance, one child expressed that item 2 was difficult for her/him because the word *website* was not familiar to him. Then, the word *website* changed as *internet sitesi* for the main study. Cronbach's Alpha coefficient of the data was found as .75. Corrected item-total correlations were between .28 (item4) to .60 (item12) which were acceptable values.

Reliability Analysis of the Final Scale

The reliability analysis of the final form of the 12-item scale pointed out that Cronbach's Alpha coefficient was .83, which implied the data had good internal consistency. Table 1 showed that the corrected-item total correlation of each item was higher than .30, which means that there were no problematic items in terms of item discrimination.

Table 1. Corrected Item-Total Correlations of Final Study

Item	Corrected Item-Total Correlation	Cronbach's Alpha If Item Deleted
Item 1	.51	.81
Item 2	.49	.82
Item 3	.49	.82
Item 4	.37	.83
Item 5	.50	.82
Item 6	.43	.82
Item 7	.52	.81
Item 8	.50	.82
Item 9	.57	.81
Item 10	.49	.82
Item 11	.52	.81
Item 12	.48	.82

Confirmatory Factor Analysis

The original scale was shown to have a one-factor structure by the scale developers. Therefore, in the CFA, the adapted version of the scale was hypothesized to have a one-factor structure. The assumptions of multivariate normality were tested by drawing a histogram and estimating skewness and kurtosis. As histogram, and skewness (-.28) and kurtosis (-.30) values implied, the data were distributed normally. There was no outlier in the data. The ratio of sample size to the number of the variable was 27.5, which implied that the sample size was sufficient. The ratio of 1 to 10 is considered as enough sample size (Bentler & Chou, 1987). The fit statistics obtained through CFA was not acceptable for the one-factor model as shown in Table 2 (CFI = .890 < .950; TLI = .866 < .950; RMSEA = .117 > .060).

Table 2. One-Factor Confirmatory Factor Analysis

	χ^2	df	χ^2 / df	CFI	TLI	RMSEA
Model 1	295.946	54	5.480	0.890	0.866	0.117

Hence, exploratory factor analysis (EFA) was conducted to understand the structure of the Turkish version. Principal axis factoring (PAF) with oblimin rotation was performed for the EFA. Kaiser-Meyer-Olkin measure of sampling adequacy value of .863 indicated that the proportion of variance in the items might be caused by the underlying factor. Bartlett's test of sphericity (p < .05) showed that the correlation matrix was different from an identity matrix. Therefore, the data was appropriate for conducting the exploratory factor analysis. As shown in Table 3, the data had a two-factor structure where items 1, 8, and 9 were loaded to a different factor.

The items that were loaded to a new factor were listed below. These three items include statements regarding mathematics, whereas the other nine items focus on science, technology, and engineering. Hence, the primary factor was called self-efficacy related to science-technology-engineering (STE), and the second factor was called self-efficacy for mathematics (Math). Items loaded to the second factor are listed below.

Item 1: "I can do math problems I get in class."

Item 8: "I think I am very good at Explaining my solutions to math problems."

Item 9: "I think I am very good at: Solving problems"

Table 3. Exploratory Factor Analysis

Item	Factor	
	1	2
Item 10	**.649**	.066
Item 11	**.633**	.022
Item 5	**.585**	.008
Item 6	**.564**	.067
Item 12	**.508**	-.060
Item 4	**.454**	.032
Item 3	**.434**	-.163
Item 2	**.428**	-.153
Item 7	**.416**	-.234
Item 8	-.034	**-.785**
Item 1	-.016	**-.776**
Item 9	.143	**-.653**

As the data structure in PAF suggested a two-factor structure, a CFA with two factors was reconducted. The two-factor model improved the fit statistics impressively as shown in Table 4 (CFI = .974 > .950; TLI = .968 > .950; RMSEA = .057 < .060). This finding showed that the STEM Competency Beliefs

scale had the two-factor structure for the Turkish data as science-technology-engineering is the first factor, and mathematics is the second factor.

Table 4. Two-Factor Analysis

	χ^2	df	χ^2 / df	p	CFI	TLI	RMSEA
Model 2	109.466	53	2.065	0.000	.974	.968	.057

Measurement Invariance

Configural, metric and scalar invariance of the scale across gender groups and career choice groups were evaluated (See Table 5 and 6). For school type, as there were a limited number of students in one group (24 students in from private school), measurement invariance analysis could not be achieved. Configural invariance results across gender groups indicated that the fit indices were good (TLI = .971, CFI = .975, RMSEA = .058). This means that the factor structure of the scale was similar for boys and girls. Metric invariance analysis showed that the change in the fit statistics supported the invariance (ΔCFI = .001, ΔRMSEA = -.003). Having metric invariance means that in addition to the factor structure, the factor loadings were equivalent across gender groups. Scalar invariance results showed that the change in the CFI was higher than allowed, whereas, for RMSEA, the change was within an acceptable range (ΔCFI = -.016, ΔRMSEA = .006). Modification indices suggested that this problem could be due to item 7. Freeing thresholds of item 7 for boys and girls resulted in better and accepted change in fit statistics (ΔCFI = -.010, ΔRMSEA = .002). This finding means that except item 7, item thresholds were invariant, and mean scores of males and females were comparable. Item 7 is "I think I am very good at: Giving evidence when I tell my opinion." Therefore, partial scalar invariance was supported for gender groups.

Configural invariance results across career choice groups indicated that fit indices were good (TLI = .961, CFI = .969, RMSEA = .063). This means that the factor structure of the scale was similar for students who want to follow STEM-related or not STEM-related careers. Metric invariance analysis showed that the change in the fit statistics supported the invariance (ΔCFI = .002, ΔRMSEA = .005). Having metric invariance means that besides the factor structure, the factor loadings were equivalent across career choice groups. Scalar invariance results showed that the changes in the CFI and RMSEA were also within acceptable ranges (ΔCFI = .000, ΔRMSEA = .009). This finding suggested that the mean scores of career choice groups are comparable.

Table 5. Measurement Invariance Analysis of the Scale for Gender Groups

	χ^2	*df*	χ^2 / *df*	TLI	CFI	RMSEA	ΔCFI	ΔRMSEA
Configural	164.13	106	1.55	.967	.974	.058 (.040; .075)	-	-
Metric	172.32	116	1.49	.971	.975	.055 (.036; .074)	.001	-.003
Scalar	230.88	138	1.67	.960	.958	.064 (.049; .079)	-.016	.006
Scalar new	215.41	135	1.60	.965	.964	.060 (.045; .075)	-.010	.002

Note. χ^2 = Chi-square, df = degrees of freedom, TLI = Tucker Lewis index, CFI = comparative fit index, RMSEA = root mean square error of approximation; CI = confidence interval, ΔCFI = change in values of CFI, ΔRMSEA = change in values of RMSEA. Scalar new: Thresholds of items 7 is freed.

Table 6. Measurement Invariance Analysis of the Scale for Career Choices

	χ^2	df	χ^2 / df	TLI	CFI	RMSEA	ΔCFI	ΔRMSEA
Configural	173.68	106	1.64	.961	.969	.063 (.045; .079)	-	-
Metric	178.68	116	1.54	.967	.971	.058 (.040; .074)	.002	.005
Scalar	203.80	138	1.48	.971	.969	.054 (.037; .069)	.000	.009

Note. χ^2 = Chi-square, df = degrees of freedom, TLI = Tucker Lewis index, CFI = comparative fit index, RMSEA = root mean square error of approximation; CI = confidence interval, ΔCFI = change in values of CFI, ΔRMSEA = change in values of RMSEA.

Comparative Analyses

Comparative analyses were conducted to test mean score differences of related groups (gender, school type, and career choices). The scores used in these comparisons were estimated using multidimensional IRT scaling. As all subgroup scores were normally distributed, a parametric test of group comparison was chosen. For the first comparison, Science, Technology, and Engineering (STE) and Mathematics (Math) score means were compared for gender groups, excluding item 7. Table 7 shows the mean score of boys and girls for STE and Math factors. Independent sample t-test showed that the mean score difference of self-efficacy on Math for boys and girls was not statistically significant ($p > .05$; d = 0.12). A similar result was found for STE mean scores of boys and girls ($p > .05$; d = 0.21).

Table 7. Mean Scores of Gender Groups

	Gender						95% CI for Mean Difference			
	Male			Female						
	M	SD	N	M	SD	n		t	df	Cohen's d
STE	.09	.90	169	-.10	.89	157	-.38; .01	1.88	324	.12
Math	.05	.96	169	-.06	.83	157	-.31; .08	1.13	324	.21

For the second comparison, STE and math factor score means were compared for public and private schools. The mean score differences between public and private school students were statistically significant for both STE and Math, as showed in Table 8. Levene's test for equality of variances indicated that the variances were equal (p =.35 for STE and p =.07 for Math). In order to assess the magnitude of the differences, effect sizes were calculated (d = 0.83 for STE, and d = 1.27 for Math). The differences between public and private school groups were significant, with large effect sizes for both STE and math (Cohen, 1988).

Table 8. Mean Differences in School Type

	School Type						95% CI for Mean Difference			
	Public			Private						
	M	SD	N	M	SD	n		t	df	Cohen's d
STE	-.05	.90	302	.64	.75	24	-1.07; -.32	-3.68***	324	.83
Math	-.07	.89	302	.88	.57	24	-1.31; -.59	-5.18***	324	1.27

*$p < .05$, **$p < .01$, ***$p < .001$

As the third comparison, the mean scores of students according to their career choices (STEM-related vs. not STEM-related) were compared. Table 9 demonstrates that there are statistically significant differences between the groups. Cohen's d was calculated for the group and obtained 0.38 for STE and 0.41 for Math. It shows the group mean scores are not equal, and they have a medium effect size.

Table 9. Mean Differences on Career Choices

	Career Choices						95% CI for Mean Difference			
	STEM Related			Not-STEM Related						
	M	SD	N	M	SD	n		t	df	Cohen's d
STE	.17	.93	161	-.17	.85	165	-.54; -.15	-3.46**	324	.38
Math	.18	.87	161	-.18	.90	165	-.55; -.16	-3.64***	324	.41

*$p < .05$, **$p < .01$, ***$p < .001$

DISCUSSION and CONCLUSION

This study contributes to the literature by adapting the STEM Competency Beliefs scale to the Turkish. Providing evidence regarding reliability and the validity of the adapted STEM Competency Beliefs

scale is expected to enable scholars to use the scale in the Turkish context. Providing measurement invariance results before comparing mean scores of scales for subgroups is also important to exemplify the procedure in comparative studies. In this respect, this study fills a gap by providing an adapted version of the newly emerging Stem Competency Beliefs scale.

An important difference between the English original and Turkish adapted scale emerged in the dimensionality of the scale. While the original scale was reported to have a one-factor structure, the Turkish scale was shown to have a two-factor structure. Item 1, 8, and 9 were loaded to a different factor, which was closely related to Math-related self-efficacy. The rest of the items were related to science, technology, and engineering. Cannady stated that the scale was also adapted into different languages as Spanish and African (M. Cannady, personal communication, November 12, 2018), and those data also showed a unidimensional structure. It can be argued that there is a sharp distinction in STEM perceptions of Turkish students as considering math in one group, and science, technology, and engineering projects in the other group. This distinction is not an expected interdisciplinary view proposed by the STEM theory. The reason for this distinction could be that Turkey does not have a direct STEM action plan, whereas many countries have a concrete strategy plan and action (MEB, 2016). Hence, students in Turkey have difficulty in perceiving STEM as a whole. Besides that, in the latest revisions of the curriculum in Turkey, there is a statement emphasizing the "science, technology, engineering" in one hand, and mathematics on the other hand (MEB, 2018a, 2018b). This might be one of the plausible explanations of why students consider STEM fields in two distinct groups. Also, studies in Turkey supported the idea that STEM is not taught in an integrative way in the schools (Baran Canbazoglu-Bilici, Mesutoglu, & Ocak, 2016; Colakoglu, 2016; Ercan, Altan, Taştan, & Dağ, 2016; Han, Yalvac, Capraro, & Capraro, 2015). All the issues mentioned here may lead students not to comprehend STEM in the actual manner.

As the mean scores of boys and girls are compared frequently throughout the scales, providing evidence regarding measurement invariance is important to get valid inferences. The measurement invariance findings showed that configural and metric invariance was supported whereas scalar invariance could be achieved freeing item 7 across gender groups. This means that the factor structure of the scale and the factor loadings were similar for boys and girls. Except for item7, threshold values to endorse statements were also similar. Therefore, excluding item 7, mean scores of boys and girls on these factors are comparable. Item 7 is related to giving evidence about opinions. This finding implies that for boys and girls, providing evidence for their opinions could have a different meaning. Similarly, measurement invariance results for student groups according to their career choices (STEM-related vs. not STEM-related) suggested that the mean scores of career choice groups could be comparable.

Comparative analysis results showed that the mean score difference of self-efficacy on Math for boys and girls was not statistically significant, as well as STE mean scores. The effect sizes also supported these findings. On the contrary to the literature (Hackett & Betz, 1982; Tellhed et al., 2017; Zeldin et al., 2008), no major differences were observed between mean scores of both STE and Math factors in Turkey. The studies in the literature generally were related to high school or older students. Hence the lower ages of the participants of this study might be an explanation for a different pattern of the findings in Turkey. It can be stated that female students are as comfortable as male students towards STEM fields in Turkey.

Secondly, it was found that students at private schools had higher self-efficacy towards STEM compared to students at public schools. This finding might be related to learning opportunities, teachers' professional development, and class size differences between school types. Many private schools promote STEM education, have STEM laboratories, and invest in robotics and technology competitions at the national and international levels. These activities and opportunities may have a positive influence on private school students. This finding is also consistent with the literature (Chittum, Jones, Akalin, & Schram, 2017; John, Bettye, Ezra, & Robert, 2016; Monterastelli, Bayles, & Ross, 2008). Additionally, teacher-related variables are an important predictor for students' academic performance (Corlu, Capraro, & Capraro, 2014). Teachers working in private schools have more opportunities to take STEM-related professional in-service training. On the other hand, public

school students mostly depend on the individual efforts of their teachers. Lastly, class size might be an explanation for the differences because private schools have smaller class sizes than public schools. Other significant differences in the scale scores were found between students who want a STEM-related career and who do not want a STEM-related career. It was observed that students who want to follow STEM-related careers had higher self-efficacy beliefs on STEM. Having an interest in STEM fields as a future career might affect these students' self-efficacy in STEM fields.

Finding significant differences between private and public school students' mean scores and between mean scores of students who want a STEM-related career or not strengthen the validity of the scale. This scale could differentiate scale scores of students who have better opportunities in private schools and who have limited resources in public schools in terms of STEM education. Additionally, this scale could assign different scores for students who want to pursue a career in STEM-related fields and for students who are not willing to pursue such a career. These findings are additional evidence for the validity of the scale (Sireci & Sukin, 2013). Therefore, this reliable and valid scale is expected to contribute to the STEM self-efficacy research in the Turkish context.

Limitations

The main limitation of the study was related to the sampling procedure. As convenience sampling was used, the generalizability of the findings could be limited. Testing the structure of the scale with another sample would provide additional evidence regarding the structure.

REFERENCES

Akgündüz, D., Aydeniz, M., Çakmakçı, G., Çavaş, B., Çorlu, M. S., Öner, T., & Özdemir, S. (2015). *STEM eğitimi Türkiye raporu*. Istanbul: Scala Publication.

Atkinson, R., & Mayo, M. (2010, December). *Refueling the U.S. innovation economy fresh approaches to science, technology, engineering and mathematics (STEM) education*. The Information Technology & Innovation Foundation, Forthcoming. Retrieved from https://ssrn.com/abstract=1722822

Bandura, A. (1999). Social cognitive theory: An agentic perspective. *Asian Journal of Social Psychology, 2*(1), 21-41.

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

Baran, E., Canbazoglu-Bilici, S., Mesutoglu, C. & Ocak, C. (2016). Moving STEM beyond schools: Students' perceptions about an out-of-school STEM education program. *International Journal of Education in Mathematics, Science and Technology*, *4*(1), 9-19. doi: 10.18404/ijemst.71338

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural equation modeling. *Sociological Methods and Research*, *16*(1), 78-117. doi: 10.1177/0049124187016001004

Bouffard-Bouchard, T., Parent, S., & Larivee, S. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school-age students. *International Journal of Behavioral Development*, *14*(2), 153-164. doi: 10.1177/016502549101400203

Breiner, J. M., Harkness, S. S., Johnson, C. C., & Koehler, C. M. (2012). What is STEM? A discussion about conceptions of STEM in education and partnerships. *School Science and Mathematics, 112*(1), 3-11. doi: 10.1111/j.1949-8594.2011.00109.x

Cai, L., Thissen, D., & du Toit, S. H. C. (2017). IRTPRO 4.2 for Windows [Computer software]. Skokie, IL: Scientific Software International.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modelling*, *14*(3), 464-504. doi: 10.1080/10705510701301834

Chen, Y. F., Cannady, M. A., Schunn, C., & Dorph, R. (2017) *Measures technical brief: Competency beliefs in STEM*. Retrieved from: http://activationlab.org/wp-content/uploads/2018/03/CompetencyBeliefs_STEM-Report_20170403.pdf

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. doi: 10.1207/S15328007SEM0902_5

Chittum, J. R., Jones, B. D., Akalin, S., & Schram, Á. B. (2017). The effects of an afterschool STEM program on students' motivation and engagement. *International Journal of STEM education*, *4*(11), 1-16. doi: 10.1186/s40594-017-0065-4

Chung, J., Cannady, M. A., Schunn, C., Dorph, R., & Vincent-Ruz, P. (2016). *Measures technical brief: Competency beliefs in science.* Retrieved from: http://activationlab.org/wp-content/uploads/2018/03/Competency-Beliefs-Report-3.2-20160331.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge.

Colakoglu, M. H. (2016). STEM applications in Turkish science high schools. *Journal of Education in Science, Environment and Health (JESEH), 2*(2), 176-187.

Colakoğlu, M. H., & Gökben, A. G. (2017). Türkiye'de eğitim fakültelerinde fetemm (stem) çalışmaları. *İnformal Ortamlarda Araştırmalar Dergisi*, *2*(2), 46-69.

Corlu, M. S., Capraro, R. M., & Capraro, M. M. (2014). Introducing STEM education: Implications for educating our teachers in the age of innovation. *Eğitim ve Bilim*, *39*(171), 74-85.

Dawes, M. E., Horan, J. J., & Hackett, G. (2000). Experimental evaluation of self-efficacy treatment on technical/scientific career outcomes. *British Journal of Guidance & Counselling*, *28*(1), 87-99. doi: 10.1080/030698800109637

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology*, *98*(2), 382-393. doi: 10.1037/0022-0663.98.2.382

English, L. D. (2016). STEM education K-12: Perspectives on integration. *International Journal of STEM Education*, *3*(3), 1-8. doi: 10.1186/s40594-016-0036-1

Ercan, S., Altan, E. B., Taştan, B., & Dağ, İ. (2016). Integrating GIS into science classes to handle STEM education. *Journal of Turkish Science Education*, *13*, 30-43. doi: 10.12973/tused.10169a

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Gainor, K. A. (2006). Twenty-five years of self-efficacy in career assessment and practice. *Journal of Career Assessment*, *14*(1), 161-178. doi: 10.1177/1069072705282435

George, D., & Mallery, P. (2001). *SPSS for Windows Step by Step: A Simple Guide and Reference*. Boston: Allyn & Bacon.

Green, A., & Sanderson, D. (2018). The roots of STEM achievement: An analysis of persistence and attainment in STEM majors. *The American Economist*, *63*(1), 79-93. doi: 10.1177/0569434517721770

Hacıoğlu, Y., Yamak, H., & Kavak, N. (2016). Mühendislik tasarim temelli fen egitimi ile ilgili ögretmen görüsleri. *Bartin Üniversitesi Egitim Fakültesi Dergisi*, *5*(3), 807-830. doi: 10.14686/buefad.v5i3.5000195411

Hackett, G., & Betz, N. (1982, March). *Mathematics self-efficacy expectations, math performance, and the consideration of math-related majors*. Paper presented at the annual meeting of the American Educational Research Association. New York.

Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on: Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, *12*(3), 38-47. doi: 10.1111/j.1745-3992.1993.tb00543.x

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2012). *Adapting educational and psychological tests for cross-cultural assessment*. New York, NY: Psychology Press.

Han, S., Capraro, R. M., & Capraro, M. M. (2016). How science, technology, engineering, and mathematics project-based learning affects high-need students in the US. *Learning and Individual Differences*, *51*, 157-166. doi: 10.1016/j.lindif.2016.08.045

Han, S., Yalvac, B., Capraro, R. M., & Capraro, M. M. (2015). In-service teachers' implementation and understanding of STEM project-based learning. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(1), 63-76. doi: 10.12973/eurasia.2015.1306a

Hidi, S., & Ainley, M. (2008). Interest and self-regulation: Relationships between two variables that influence learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 77-109). Lawrence Erlbaum Associates Publishers.

International Test Commission. (2017). The ITC guidelines for translating and adapting tests (2nd ed.). Retrieved from https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf

Jenson, R. J., Petri, A. N., Day, A. D., Truman, K. Z., & Duffy, K. (2011). Perceptions of self-efficacy among STEM students with disabilities. *Journal of Postsecondary Education and Disability*, *24*(4), 269-283.

Jinks, J., & Lorsbach, A. (2003). Introduction: Motivation and self-efficacy belief. *Reading and Writing Quarterly, 19*(2), 113-118. doi: 10.1080/10573560308218

John, M., Bettye, S., Ezra, T., & Robert, W. (2016). A formative evaluation of a southeast high school integrative science, technology, engineering, and mathematics (STEM) academy. *Technology in Society*, *45*, 34-39. doi: 10.1016/j.techsoc.2016.02.001

Johnson, C. C., Peters-Burton, E. E., & Moore, T. J. (2016). *STEM road map: A framework for integrated STEM education*. New York, NY: Routledge, Taylor & Francis Group.

Kanny, M. A., Sax, L. J., & Riggers-Piehl, T. A. (2014). Investigating forty years of STEM research: How explanations for the gender gap have evolved over time. *Journal of Women and Minorities in Science and Engineering*, *20*(2), 127-148. doi: 10.1615/JWomenMinorScienEng.2014007246

Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of Counseling Psychology*, *33*(3), 265-269. doi: 10.1037/0022-0167.33.3.265

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in crosscultural research. *International Journal of Psychological Research, 3*(1), 111-121.

Milner, D. I., Horan, J. J., & Tracey, T. J. (2014). Development and evaluation of STEM interest and self-efficacy tests. *Journal of Career Assessment, 22*(4), 642-653. doi: 10.1177/1069072713515427

Ministry of National Education. (2009). *Ortaöğretim okulları öğrenci kulüp faliyetlerine yönelik eğitim materyali ve donanım ihtiyacının değerlendirilmesi*. Ankara: EARGED.

Ministry of National Education. (2016). *STEM Education Report*. Ankara: SESAM.

Ministry of National Education. (2018a). *Fen bilimleri dersi öğretim programı*. Ankara: MEB.

Ministry of National Education. (2018b). *Matematik dersi öğretim programı*. Ankara: MEB.

Monterastelli, T., Bayles, T., & Ross, J. (2008, June). *High school outreach program: Attracting young ladies with "engineering in health care."* Paper presented at the Annual Conference, & Exposition, Pittsburgh, Pennsylvania. Retrieved from https://peer.asee.org/3621

Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén and Muthén.

National Research Council. (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. Washington, DC: National Academic Press.

Nelson, B. C., & Ketelhut, D. J. (2008). Exploring embedded guidance and self-efficacy in educational multi-user virtual environments. *Computer-Supported Collaborative Learning, 3*(4), 413-427. doi: 10.1007/s11412-008-9049-1

Pajares, F., & Miller, M. D. (1997). Mathematics self-efficacy and mathematical problem solving: Implications of using different forms of assessment. *The Journal of Experimental Education, 65*(3), 213-228. doi: 10.1080/00220973.1997.9943455

Reckase, M. D. (2009). *Multidimensional item response theory.* New York, NY: Springer. doi: 10.1007/978-0-387-89976-3

Sireci, S. G., & Sukin, T. (2013). Test validity. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbooks in psychology. APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 61-84). Washington, DC: American Psychological Association. doi: 10.1037/14047-004

Smith, S. M. (2019). *A comparison of computer-based and robotic programming instruction: Impact of Scratch versus Cozmo on middle school students' computational thinking, spatial skills, competency beliefs, and engagement* (Doctoral dissertation, Kent State University).

Tellhed, U., Backström, M., & Björklund, F. (2017). Will I fit in and do well? The importance of social belongingness and self-efficacy for explaining gender differences in interest in STEM and HEED majors. *Sex Roles, 77*(1-2), 86-96. doi: 10.1007/s11199-016-0694-y

Tsupros, N., R. Kohler, and J. Hallinen (2009). *STEM education: A project to identify the missing components*. Intermediate Unit 1 and Carnegie Mellon, Pennsylvania.

TÜSİAD (2019). *2023'e doğru Türkiye'de STEM gereksinimi*. https://tusiad.org/tr/yayinlar/raporlar/item/9735-2023-e-dog-ru-tu-rkiye-de-stem-gereksinimi adresinden edinilmiştir.

Ullman, J. B. (2001). Structural equation modeling. In B. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (4th ed.) (pp. 653-771). Boston, MA: Allyn & Bacon.

Urdan, T. C. (2010). *Statistics in plain English* (3rd ed.). New York, NY: Taylor & Francis Group.

Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*(2), 89-99.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70. doi: 10.1177/109442810031002

Vincent- Ruz, P., & Schunn, C. D. (2017). The increasingly important role of science competency beliefs for science learning in girls. *Journal of Research in Science Teaching*, *54*(6), 790-822. doi: 10.1002/tea.21387

Yerdelen, S., Kahraman, N., & Taş, Y. (2016). Low socioeconomic status students' STEM career interest in relation to gender, grade level, and STEM attitude [Special issue]. *Journal of Turkish Science Education (TUSED), 13*, 59-74.

Zeldin, A. L., Britner, S. L., & Pajares, F. (2008). A comparative study of the self- efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching*, *45*(9), 1036-1058. doi: 10.1002/tea.20195

Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura, (Ed.), *Self-efficacy in changing societies* (pp. 202-231). New York, NY: Cambridge University Press.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*(1), 82-91. doi: 10.1006/ceps.1999.1016

Bilim, Teknoloji, Matematik ve Mühendislik Alanlarında Öz Yeterlik İnanç Ölçeğinin Türkçe'ye Uyarlaması ve Ölçme Değişmezliğinin Test Edilmesi

Giriş

Bilim, Teknoloji, Mühendislik ve Matematik alanları (STEM) eğitimi bu alanların bir bütün olarak ele alınması ile günlük yaşam problemlerinin çözümü ile ilgilenmektedir (Breiner, Harkness, Johnson, & Koehler, 2012; Johnson, Peters-Burton, & Moore, 2016; National Research Council-NRC, 2014; Tsupros, Kohler, & Hallinen, 2009). Türkiye de STEM eğitimine önem veren ve bu konuda yatırım yapan ülkelerdendir (Akgündüz ve diğerleri, 2015; TÜSİAD, 2019). Öz yeterlik inançları akademik başarıda önemli rol oynayan faktörlerden birisidir (Kanny, Sax & Riggers-Piehl, 2014). Ayrıca araştırmalar öz yeterlik inançları ve ilgi arasında pozitif bir ilişki göstermektedir (Hidi & Ainley, 2008). Bunun yanı sıra bireyler mesleki tercihlerini yaparken başarılı olacaklarını düşündükleri alanları tercih etmektedirler (Durik, Vida, & Eccles 2006; Gainor, 2006). Bu sebeple STEM eğitimi çerçevesinde öğrencilerin öz yeterlik inançlarını ölçerek STEM eğitimi ile ilişkilendirmek önemlidir. Ancak, Türkiye'de STEM öz yeterlik becerilerini ile ilgili bir ölçek bulunmamaktadır. Chen, Cannady, Schunn ve Dorph (2017) İngilizce olarak STEM yeterlik inançları ölçeği geliştirmiştir. Bu çalışma da bu ölçeğin Türkçe'ye uyarlamasını yapmayı amaçlamaktadır. Bu ölçeğin Türkçe'ye kazandırılmasının Türkiye'deki STEM çalışmalarına katkı sağlaması beklenmektedir. Bu çalışmanın iki ana amacı bulunmaktadır. Birinci amaç ölçeğin uyarlanarak yapısının Türk öğrencilerden toplanan veri ile test edilmesidir. İkinci amaç ise yapının cinsiyet grupları ve STEM ile ilgili kariyer hedefi olan ve olmayan öğrenci grupları arasında ölçme değişmezliği gösterip göstermediğinin incelenmesidir. Ayrıca öğrencilerin ölçekte elde edilen puanları cinsiyet, okul türü ve kariyer hedefleri değişkenleri bakımından karşılaştırılmıştır. Bu çalışmanın araştırma soruları aşağıdaki gibidir.

1) STEM yeterlik inançları ölçeğinin orijinal yapısı Türk öğrencilerinin verisi ile desteklenmekte midir?
2) Elde edilen yapı kızlar ve erkekler için ölçme değişmezliği göstermekte midir?
3) Elde edilen yapı STEM ile ilgili kariyer hedefi olan ve olmayan öğrenciler için ölçme değişmezliği göstermekte midir?
4) Öğrencilerin ölçekte elde edilen puanları cinsiyet, okul türü ve kariyer hedefleri değişkenleri bakımından farklılık göstermekte midir?

Yöntem

Örneklem

Uyarlama aşaması pilot uygulama ve asıl uygulama basamaklarından oluşmuştur. Pilot uygulamaya 77 ortaokul öğrencisi, asıl uygulamaya 330 ortaokul öğrencisi katılmıştır. Asıl uygulamada kız ve erkek sayıları birbirine yakındır. Öğrencilerin %92'si devlet okulu, %8'i ise özel okul öğrencisidir.

Ölçme aracı

Ortaokul öğrencilerinin STEM yeterlik inançlarını ölçmeyi amaçlayan bu ölçek 12 maddeden oluşmakta ve 4'lü Likert tipi yapıya sahiptir. Ölçme aracı "Sınıfta sorulan matematik sorularını çözebilirim" ve "Evimdeki teknoloji uzmanı benim" gibi maddelerden oluşmaktadır.

Veri analizi

Veri analizi kısmında uyarlama aşamaları, pilot çalışma, güvenirlik ve geçerlik analizleri ve ölçme değişmezliği analizleri ile ilgili yapılanlar açıklanmaktadır.

Uyarlama aşamasında ilk olarak gerekli izinler alınmıştır. Ardından bu konuda tecrübeli uzmanlar tarafından ölçeğin çevirisi gerçekleştirilmiştir. Bağımsız yapılan bu çeviri işleminden sonra araştırma ekibinin de sürece dahil olması ile bu aşama tamamlanmıştır. Ardından geri çeviri aşaması gerçekleştirilmiştir. Son aşama olarak ölçeğin Türkçesi uzmanlar tarafından incelenmiştir. Araştırma ekibi ise gerekli kontrolleri yapmıştır. Pilot aşamasında ifadelerin anlaşılırlığı incelenmiş ve gerekli düzeltmeler yapılmıştır.

Güvenirlik için Cronbach Alfa iç tutarlılık katsayısı hesaplanmıştır. Bu değerin .70'ten büyük olması beklenmektedir. Ayrıca, madde bazında sorunları görebilmek için düzeltilmiş madde toplam korelasyonu hesaplanarak değeri .30 altında olan maddeler incelemeye alınmıştır.

Geçerlik çalışmaları için doğrulayıcı faktör analizi yapılmıştır. Doğrulayıcı faktör analizinde daha önce belirlenmiş olan bir yapının toplanan verilerle uyumu incelenir. CFI (Comparative Fit Index), TLI (Tucker Lewis index) ve RMSEA (Root Mean Square Error of Approximation) gibi örneklem sayısından direk etkilenmeyen uyum değerleri incelenerek testin yapısı test edilmektedir. CFI ve TLI değerlerinin .95'ten büyük, RMSEA değerinin ise .06'dan küçük olması istenmektedir (Ullman, 2001). Orijinal ölçekte belirlenen tek faktörlü yapı doğrulayıcı faktör analizi kapsamında test edilmiştir.

Gruplar arası karşılaştırma akademik çalışmalarda önemli bir yer tutmaktadır. Ancak bu karşılaştırmaların yapılabilmesi için ölçülen kavramların alt gruplar için aynı anlam taşıyıp taşımadığı test edilmelidir. Bu sebeple yapısal, metrik ve skalar değişmezlik incelenmiştir. Yapısal modelde gruplar için yapı benzerliğine, metrik modelde faktör yüklerinin eşitliğine, skalar modelde ise ortalamaların eşitliğine bakılmıştır. Modeller arası uyum değeri farkının CFI için .01'den RMSEA için .015'ten küçük olması istenir (Chen, 2007; Cheung & Rensvold, 2002). Öğrenci puanları ise çok boyutlu madde tepki kuramı kullanılarak kestirilmiştir.

Sonuç ve Tartışma

İç tutarlılık

12 maddeden oluşan ölçeğin Cronbach Alfa iç tutarlılık değeri .83 olarak hesaplanmıştır. Bu değer ölçeğin iyi düzeyde iç tutarlılığa sahip olduğunu göstermektedir. Düzeltilmiş madde-toplam korelasyon değerlerinin hepsinin .30 değerinin üzerinde olması ise madde bazında bir problem olmadığını göstermektedir.

Doğrulayıcı faktör analizi

Doğrulayıcı Faktör Analizi (DFA) sonuçlarına göre elde edilen veri tek faktörlü yapıyı desteklememektedir (CFI = .890 < .95; TLI = .866 < .95; RMSEA = .117 > .06). Bu sebeple Açımlayıcı Faktör Analizi (AFA) yapılarak faktör yapısı incelenmiştir. AFA sonuçları ölçekteki maddelerin 2 farklı boyut oluşturduklarını göstermektedir. Madde 1, 8 ve 9 ayrı bir faktör ile ilişkidirler. Bu maddeler incelendiğinde bu maddelerin matematik ile ilgili oldukları diğer maddelerin ise bilim, teknoloji ve mühendislik ile ilgili oldukları görülmektedir. Bu faktörlere Mat ve STE isimleri

verilmiştir. Burada ortaya çıkan iki boyutlu yapı DFA ile incelendiğinde ise yapının veri tarafından doğrulandığı görülmektedir (CFI = .974 > .95; TLI = .968 > .95; RMSEA = .057 < .06). Bu sebeple ölçeğin Türkçe uyarlamasının iki boyutlu bir yapıya sahip olduğuna karar verilmiştir.

Ölçme değişmezliği

Kızlar ve erkeklerden elde edilen veri yapı değişmezliğini desteklemektedir (TLI = .971, CFI = .975, RMSEA = .058). Faktör yüklerinin eşitlenmesi ile elde edilen model karşılaştırması da metrik değişmezliğin olduğunu ortaya koymaktadır (ΔCFI = .001, ΔRMSEA = -.003). Ancak, skalar değişmezlik verilerinde ΔRMSEA değeri iyi iken ΔCFI değeri istenen seviyede değildir (ΔCFI = -.016, ΔRMSEA = .006). Modifikasyon değerleri bu sorunun 7. maddeden kaynaklanabileceğini göstermektedir. Bu madde üzerindeki sınırlılıklar kaldırıldığında ise elde edilen değerler skalar değişmezliğin de desteklendiğini göstermektedir (ΔCFI = -.010, ΔRMSEA = .002). Bu sebeple madde 7 dışında testin ölçme değişmezliğine sahip olduğu ve kızlar ve erkeklerin puanlarını karşılaştırmada kullanılabileceği sonucuna ulaşılmıştır. Madde 7 "Fikrimi söylerken kanıtlar sunmakta iyiyim" ifadesinden oluşmaktadır.

Kariyer hedefleri STEM ile ilgili olan ve olmayan öğrenciler için de ölçme değişmezliği test edilmiştir. Kariyer hedefi grupları için elde edilen veri yapı değişmezliğini desteklemektedir (TLI = .961, CFI = .969, RMSEA = .063). Faktör yüklerinin eşitlenmesi ile elde edilen model karşılaştırması da metrik değişmezliğin olduğunu ortaya koymaktadır (ΔCFI = .002, ΔRMSEA = .005). Skalar değişmezlik verilerinde de ΔCFI değeri ve ΔRMSEA değeri beklenen düzeydedir (ΔCFI = .000, ΔRMSEA = .009). Bu bulgular kariyer grup ortalamalarının karşılaştırılabileceğini göstermektedir.

Grupların karşılaştırılması

Öğrencilerin çok boyutlu madde tepki kuramı kullanılarak kestirilen mat ve STE puanları cinsiyet, okul türü ve kariyer hedefleri değişkenleri bakımından karşılaştırılmıştır. Kızlar ve erkekler arasında istatistiksel olarak anlamlı bir fark bulunamamıştır. Özel okullardaki öğrencilerin devlet okullarındaki öğrencilere göre öz yeterlik inanç puanlarının daha yüksek olduğu görülmüştür. İleride STEM ile ilgili alanlarda bir meslek sahibi olmak isteyen öğrencilerin puanları STEM dışında mesleklere yönelmek isteyen öğrencilerin puanlarından daha yüksektir. Bu sonuçlar etki büyüklüğü hesapları tarafından da doğrulanmaktadır.

Bu çalışma STEM öz yeterlik inançları ölçeğini Türkçe'ye uyarlaması bakımından önemli bir çalışmadır. Ölçeğin güvenirliği ve geçerliği ile ilgili kanıtlar sunulmuş, STEM araştırmalarında kullanılabilecek bir uyarlama olduğu ortaya konmuştur. Karşılaştırma çalışmalarında bir önkoşul olan ölçme değişmezliğinin test edilmesi ve örneklendirilmesi de önemlidir.

Elde edilen veriler uyarlanan ölçeğin faktör yapısının orijinal ölçeğin faktör yapısından farklı olduğunu göstermiştir. Bu durumun Türkiye'de STEM kavramlarının bir bütün olarak görülmemesinden kaynaklandığı düşünülmektedir. Öğretim programlarındaki vurgunun da bir bütün oluşturmadığı görülmektedir.

Özel okullardaki öğrencilerin ve STEM ile ilgili bir kariyer isteyen öğrencilerin daha yüksek STEM öz yeterlik inanç puanına sahip olmaları geçerlik için ayrıca bir kanıt olduğu düşünülmektedir. Bu ölçeğin puanları farklı öğrenci grupları için farklılık gösterebilmektedir. Özel okullarda sağlanan STEM imkanları ve uygulamaları ile devlet okullarının kısıtlı imkanları öğrencilerin öz yeterliklerinin ayrışmasına sebep olmuş olabilir. STEM ile ilgili kariyer hedefleyen öğrenciler ile farklı alanlara yönelmek isteyen öğrencilerin öz yeterlik puanlarının farklı çıkması da bu ölçeğin geçerliğini desteklemektedir.