International Journal of Psychology and Educational Studies, 2020, 7(2), 111-121

PES

www.ijpes.com

International Journal of Psychology and Educational Studies

A Validity and Reliability Study of the Basic STEM Skill Levels Perception Scale

Özgen Korkmaz[1], **Recep Çakır**[2], **Feray Ugur-Erdogmus**[3]

[1]*Amasya University, Technology Faculty, Department of Computer Engineering, Amasya, Turkey*

[2]*Amasya University, Education Faculty, Department of CEIT, Amasya, Turkey*

[3]*Vocational School of Design, Computer Aided Design and Animation, Amasya, Turkey*

ARTICLE INFO

Article History:
Received 13.01.2020
Received in revised form 20.02.2020
Accepted 27.02.2020
Available online 04.05.2020

ABSTRACT

The aim of this study is to develop a perception scale related to the possible basic skills that can be gained through STEM. Participants of this study were 723 university students. In this study, descriptive survey study was conducted. To identify validity of the scale exploratory factor analysis, cumulative item factor, corrected correlations and item discrimination were calculated. For reliability internal consistency and stability level were calculated. Collected data were analyzed in terms of arithmetic mean, standard deviation, t and ANOVA. The scale is a 7-point likert-type scale which consists of 43 items under 3 factors. Data analysis results showed that this scale is valid and reliable for measuring students' STEM skills according to their perceptions.

Keywords:[1]
Stem Skills; Scale Development; Validity; Reliability

1. Introduction

Along with the developing world economy, science and technology have improved further since the second half of 20th century (Yıldırım & Selvi, 2017). Considering this realm, training productive learners who are inquisitive about science; who are problem solvers and creative, self-perpetuating individuals has become one of the most important objectives of education. To develop 21st century skills and train these individuals, countries should review and reform their educational systems (Bybee, 2013). As it is the case for many subject areas, when a glance is taken at the aims of science educational program, training students as science literate individuals is of crucial importance. It is expected from the science literate individuals to be collaborative, to make good use of communication skills, to be a life-long learner and to embrace knowledge, skills, attitudes, perceptions and values that science contains (Stinson, Harkness, Meyer & Stallworth, 2009). On the other hand, science literate individuals can use scientific knowledge, interpret world by defining problems based on evidences and derived results, and determine what kind of changes may occur and their causes according to possible changes of human activities (Rogers & Porstmore, 2004). Individuals want to explore and transfer knowledge through the generations by developing these explorations and inventions. These explorations and inventions are attained by converting theoretical knowledge into usable forms (Daugherty, 2009; Yılmaz, Gülgün & Çağlar, 2017). For sure, generating theories, developing explorations and inventions based on these theories require an interdisciplinary effort.

Science, Technology, Engineering and Mathematics (STEM) education aims to approach problems with an interdisciplinary approach (Honey, Pearson & Schweingruber, 2014; Yılmaz, Gülgün & Çağlar, 2017). STEM

[1] Corresponding author's address: Amasya University, Technology Faculty, Department of Computer Engineering, Amasya, Turkey
Telephone: +90358 260 00 66 – 1515 ,
e-mail: ozgenkorkmaz@gmail.com
http://dx.doi.org/10.17220/ijpes.2020.02.010

Education focuses on three themes: gaining problem solving skills, being innovative and being capable of designing (Hernandez, et. al., 2014; Yılmaz, Gülgün & Çağlar, 2017). Problem solving-oriented interdisciplinary point of view is accompanied with teaching Science, Technology, Engineering and Mathematics areas in a combined fashion and this point of view enabled the formation of STEM education (Hernandez, et al., 2014, 108; Gülhan & Şahin, 2016). By definition, STEM aims to take individuals at an interdisciplinary level that includes Science, Mathematics, Engineering and Technology areas (Sullivan, 2008). Within this framework STEM is an approach that teachers support students constantly and encourage them to produce, to solve problems and to be in search of a better solution than the usual solution(s) (Yılmaz, Gülgün & Çağlar, 2017).

As the STEM education gained popularity, quality of it began to be questioned. It is suggested that to educate creative and innovative individuals should include science, technology, engineering and mathematics areas and this approach is called as STEAM (Yakman, 2008). Yakman (2008) formed a pyramid and divided it into four areas to define the STEAM education. In this pyramid at the lowest step topics of the STEAM disciplines are placed in a separate manner. The *disciplined specific* upper step is placed on the top of the first step and disciplines formed on these topics. Then at the third, *multidisciplinary* step, STEM education and art are placed. The fourth step which is called *integrative* step represents STEAM that integrates all disciplines and arts.

Before introducing STEM education, first theoretical framework should be taught (Yılmaz, Gülgün & Çağlar, 2017). This framework holds the knowledge of STEM subjects, integration and 21st century skills. 21st century skills point out upper-level thinking skills that students need for fulfilling necessities of the information age and also underline their learning tendencies (Whittle, Pell & MurdochEaton, 2010). These skills still constitute a requirement for students and schools are struggling for this very purpose (Günüç, Odabaşı & Kuzu, 2013). Students should be the individuals who are flexible, able to take initiatives if required and able to produce new and useful products (Gelen, 2007). According to World Economic Forum while skills that are going to be needed in 2015 were listed as complex problem solving, coordinating with others, people management, critical thinking, negotiation, quality control, service orientation, judgment and decision making, active listening and creativity; skills that are going to be needed in 2020 are articulated as complex problem solving, critical thinking, creativity, people management, coordinating with others, emotional intelligence, judgment and decision making, service orientation, negotiation and cognitive flexibility (Gökkurt, Örnek, Hayat & Soylu, 2015). While maintaining their academic/learning life, being aware of the skills related to the demands of their future vision, the ones for the business world is essential for learners and developing themselves in accordance with these skills is critical to that end (MNE, 2011). All skills that are expected to be even more important in the 21st century are closely related to lifelong learning concept. Sustaining lifelong learning activities to develop students' knowledge, skills and abilities with respect to personal, citizenship, social and/or employment perspectives will be a suitable approach to be able to target the necessities of this age (Keçeci, Alan & Kırbağ Zengin; 2017). According to MNE (2011) skills that students should have in the 21st century addressed in four main themes, which are ways of thinking, ways of working, working tools and integration with world. Innovative/creative thinking and to be open for these, critical thinking, problem solving and decision making, using learning strategies/learning how to learn and having higher-order cognitive skills and self-evaluation can be considered within the scope of 'the ways of thinking'.

In this regard, in the STEM education process measurement of both academic achievements and the 21st century skills of the students are necessary. It is possible to utter that performance evaluation is the most common measurement and evaluation method in STEM education. In this method there are two parts: performance task and rubrics. The difference between the performance tasks and multiple-choice tests, which require choosing the right choice, is that students produce their own answers rather than arriving at the right choice amongst the provided ones. Rubric is the most common measurement tool. It is possible to come across rubric examples in the literature that are used for identifying to what extent students have the 21st century skills. However, these rubrics are the tools which are used in a rather discipline-dependent way. Moreover, teachers should observe students one by one in rubrics. Result of the literature review showed that there is not any valid or reliable scale developed to measure, STEM skills. In addition, there are limited numbers of studies related to possible basic skills that can be gained through STEM. Therefore, the aim of this study is to develop a valid and reliable perception scale to measure basic STEM skills levels of students according to their perceptions.

2. Methodology

2.1. Research Method

This study is a scale development study as well as being a descriptive one. A survey model was carried out. In this scope basis STEM skill levels perception scale was developed.

2.2. Participants

Participants of this study consisted of 723 university students who were students of elementary mathematics education, elementary science education and computer and instructional technology departments of education faculty and electrical and electronics engineering and mechanical engineering departments of engineering faculty of Amasya University in 2018 spring semester. While administering study groups, quantitative fields are chosen because students of these fields are thought to possess basic mathematics, science, engineering and technology skills. Volunteering students participated in this study. 361 of the participants were female and 362 of them were male. Distribution of the participants according to department and grade level is summarized in Table 1.

Table 1. Distribution of participants according to grades and departments

Department	Grade				Total
	1	2	3	4	
Elementary Mathematics Education	35	46	42	34	157
Elementary Science Education	23	44	58	74	199
Computer Education and Instructional Technology	3	34	38	0	75
Electrical and Electronics Engineering	45	60	50	45	200
Mechanical Engineering	40	52	0	0	92
Total	146	236	188	153	723

It is assumed that the students who will be included in the scope of the study have sufficient knowledge in engineering, science, mathematics and technology. When the curricula of these departments are examined, it is seen that there are many courses related to basic STEM skills. Hence these departments were selected as places to collect the relevant data. In that sense it can be expected that the differences between the items on the scale will be high.

2.3. Development process of the scale

Development process of this scale was initiated with literature review and creation of an item pool. There are many achievement scales related to STEM in the literature. Because of the nature of STEM almost all of these scales are comprised of field-specific open-ended questions and rubrics (Çepni, 2018). In this study, it is intended to develop a scale related to the possible basic skills that can be gained through STEM and how these skills are measured in support of field-specific test in the literature. Within these frameworks qualified items which could be answers to the following questions were attempted to be produced. To ensure that the item pool is as rich as possible, the questions are kept fairly general.

(1) Which skills are aimed to be developed with STEM?
(2) How does a student who has STEM skills behave?
(3) What kind of actions a student makes if they possess STEM skills?
(4) What are the target learning outcomes at the end of the STEM process in general?
(5) What are the evidences that show students have STEM skills?

Moreover, rubrics and learning outcomes pertaining to different levels of curriculum presented on the web pages of Stanford NGSS Assessment Project (SNAP, 2018) and Next Generation Science Standards (NGSS, 2018) are converted to a field-independent format and added onto the following item pool. For example, a learning outcome placed in DCI Arrangements of the Next Generation Science Standards (NGSS, 2018) and stated as "3-PS2-4. Define a simple design problem that can be solved by applying scientific ideas about magnets." was adapted as a can-do statement: "I can define simple design problems that can be solved by applying scientific ideas about any subjects". Another learning outcome that is stated as "4-PS3-4. Apply scientific ideas to design, test, and refine a device that converts energy from one form to another." was adapted as "I can apply scientific ideas to design a device for solving a problem. Similarly, the learning outcome which is stated as "5-LS2-1. Develop a model to describe the movement of matter among plants, animals, decomposers, and the environment." was revised as "I can describe a topic by modelling all of its components." Moreover, the learning outcome which is stated as "MS-PS4-1. Use mathematical representations to describe a simple model for waves that includes how the amplitude of a wave is related to the energy in a wave" was adapted as "I can use mathematical representations to describe a simple model related to science". Likewise, the learning outcome that is stated as "Analyze and interpret patterns in graphs/charts/maps to make predictions about natural hazards" and was placed in the curriculum developed by SANP (2018) was converted into two items: "I can analyze a graph and explain the relationship between the quantities in it" and "I can make predictions about a phenomenon by observations". Also, the learning outcome that is stated as "Integrate quantitative or technical information expressed in words in a text with a version of that information expressed visually (e.g., in a flowchart, diagram, model, graph, or table)" was adapted as "I can express evidences that I collected for the solution of a problem in graph or table".

Resulting item pool was finalized by using Delphi technique. Traditionally, Delphi technique requires application of a set of phases to reveal point of views of domain experts of a problem and to arrive at agreement as to what the solution to that problem should be. In this study, explained phases (Şahin, 2001) are applied as it is described in the following lines:

First step of the Delphi technique is to specify the problem. In this study, the following question forms the problem statement: "Which learning outcomes should an individual have to be evaluated as being capable of basic STEM skills?". This problem is stated as five questions, which were provided above, for them to be grasped in similar ways by the domain experts. The second step is determining panel members. In this context, eight domain experts were identified. All these domain experts have PhD degrees. Four of them have their PhD degree in instructional technology, two of them in science education, one of them in measurement and evaluation and the other one in primary mathematics education field. All were working on STEM then. In the third step, five questions mentioned above were emailed to these domain experts. As part of developing the first Delphi survey, to determine as many skills as possible related to topic the domain experts were asked to brainstorm. In the second Delphi survey formation step, opinions of the domain experts were listed as items and placed under "basic science skills, basic mathematics skills, basic engineering skills and basic technology skills" subtitles. As explained above, skills that were placed in the curriculum of SNAP (2018) and NGSS (2018) were added to skills that were written by the domain experts. The acquired item pool was resent to the domain experts. They were requested to take notes their opinions and justify the reasons behind agreeing/ disagreeing on items and considering whether the items are important or not using the related margins. In the third Delphi application, the domain experts were requested to review their answers related to the second item pool. Mean scores and expert opinions related to the previous application were also added to the form, which was submitted to domain experts. The domain experts were requested to review and write their opinions according to the mean scores. After this step, a meeting was assembled with the domain experts and in this meeting each item was reviewed one by one and the item pool was finalized. The draft form consisted of 66 items. In this meeting there were four domain experts. In the draft form items were organized into a 7-point likert type scale presenting statements to "strongly disagree" to "strongly agree" with. To refrain from misunderstandings short definitions of concept, analyze, analog signal, reference, relation, inference, digital signal, model, quantitative, phenomenon, abstract, design and data were added to the draft form.

2.4. Data analysis

In the statistical analysis, first, KMO and Bartlett tests were applied on the data that was gathered with the scale to determine construct validity of it and to determine whether factor analysis can be administered or not. Based on these results an explanatory factor analysis was conducted; factor discrimination of the scale was determined by principal component analysis; Varimax orthogonal rotation was used to analyze factor loadings. Independent sample t-test was used to test the discriminating power of items that remained in the scale after the factor analysis. Validity of the scale was determined by testing total item correlation of the scale by Pearson's r test. Item discrimination was tested by observing discrimination between lower 27% and upper 27% groups. Coefficient of internal consistency and stability was conducted to measure reliability of the scale. Cronbach's alpha reliability coefficient, split-half reliability correlation, Spearman-Brown formula and

Guttmann split-half reliability formula were used to determine internal consistency level. Test-retest method was conducted to identify the scale's stability.

3. Results

Findings related to validity of the scale

For the validity of the Basic STEM Skill Levels Perception Scale, construct validity, total item correlations and item discrimination levels were calculated, and findings were represented below.

Construct validity

Tatlıdil (2002) states that first appropriateness to factor analysis of collected data should be tested. Hence, Kaiser-Meyer-Oklin (KMO) and Bartlett tests were conducted, with a view to determining if exploratory factor analysis could be conducted or not. KMO values that were between 0.70 and 0.80 were considered as middling, 0.80 to 0.90 were good and higher than 0.90 were interpreted as the data set is marvelous for factor analysis. Moreover, if KMO value is under 0.50, data set cannot be factorized (Field, 2000; Russell, 2002). Besides, according to Bartlett test null hypothesis is rejected at 0.05 significance level (Büyüköztürk, 2002; Eroğlu, 2008). In this study KMO = 0.936; Bartlett test results were χ^2= 40417.699; *df*=2145 (*p*=0.000). Hence, data set is marvelous for factor analysis. On the other hand, common factor variances are between .585 to .821. According to this result all the items can be considered quite good.

In this framework, exploratory factor analysis was conducted; principal component analysis was carried out to identify scales factors; and by conducting Varimax rotation technique factor loads were investigated. Factor analysis is used to discover whether items of a scale are distributed to less factors or not (Balcı, 2009; Carmines, 1982). Principal component analysis is a common technique for factorization (Büyüköztürk, 2002; Carmines, 1982). According to principal component analysis results, if item factor loads are below 0.40 or items whose difference between two factor loads are not at least 0.100 (i.e., items whose factors distributed on two factors) should be omitted (Büyüköztürk, 2002). In the first analysis, when natural factor distribution was examined, there were 13 factors whose eigenvalue were above 1. However, a considerable part of the items was gathered under 3 factor and eigenvalues of these factors were quite large. Thus, factor analysis started as three-factor solution. According to principal components of tree-factor structure Varimax orthogonal rotation technique was conducted and 27 items, whose item loads were under 0.40 and distributed different factors, were removed from the scale incrementally. Removed items were revised by four field experts who participated in the third Delphi round meeting. Field experts came to an agreement on that learning outcomes that were measured by 17 of the removed items were also measured by other remaining items of the scale; and 6 of the removed items were slightly not appropriate to the general aim of the scale. Although factor loads of 4 items were below 0.40, they were not removed from the scale since it was considered that it could impact content validity negatively. As a result, it was agreed that removed items would not affect the content validity negatively and factor analysis was reconducted on 43 remained items.

At the end of these processes, it was witnessed that remaining 43 items gathered under 3 factors. KMO value of 43 itemed scale was 0.947; Bartlett values of it was χ^2=23874.641 df=903; p<0.001. Main criteria for factor analysis results is factor loads (Balcı, 2009; Eroğlu, 2008; Gorsuch, 1983). High factor loads are considered as a sign of possibility of placement of the variable under respective factor (Büyüköztürk, 2002). Unrotated factor loads of remaining 43 items were between .461 to .758. In the literature, explaining at least 40% of the general variance is found adequate in behavioral sciences (Büyüköztürk, 2002; Eroğlu, 2008; Kline, 1994; Scherer at al., 1988). First, it was identified that items in the scale and factors explain 54.16% of the variance in total variance. Then, contents of the items of factors were examined and checked to see whether they were placed under the predetermined themes (Science, Mathematics, Engineering and Technology). It was observed that Engineering and Technology themes merged. In this framework factors were entitled as Science, Mathematics and Engineering-technology. Factor structure can also be seen in eigenvalue scree plot (Figure 1). In figure 1, high accelerated decline is observed in the first three factors, so these three factors have important contributions to the variance. On the other hand, other factors decline horizontally meaning that contribution of these factors to the variance are close the each other (Büyüköztürk, 2002; Eroğlu, 2008).

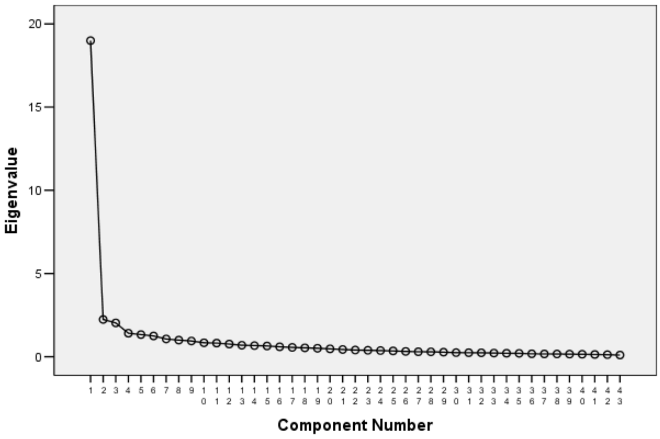

Figure 1. Screen plot graphic (eigenvalues according to the factors).

Findings that are related to item loads of remaining 43 items with respect to factors, eigenvalues of factors and the part of explained variance which is related to factors are presented in Table 2.

Table 2. Factor analysis results of the scale as per factors

		Items	Com. Var.	F1	F2	F3
Science	1.	I can form relation/relations related topics.	.701	.758		
	2.	I can form a model according to evidences that I gathered.	.653	.745		
	3.	I can make prediction according to my observations related to a phenomenon.	.607	.734		
	4.	I can make predictions by making measurements related to a phenomenon.	.516	.688		
	5.	I can analyze a problem up to the minimum components.	.593	.674		
	6.	I can assert a claim related to a topic by considering all of the evidences.	.560	.644		
	7.	I can conduct a research to prove a claim, phenomena or theory.	.563	.642		
	8.	I can compare different solutions that I produce for a solution of a problem.	.604	.618		
	9.	I can criticize others' ideas by producing scientific/realistic/concreate/evidence-based claims.	.557	.618		
	10.	I can explain relations of information that is needed for forming a model by going over a sample.	.547	.608		
	11.	I can design a basic problem in which material, time or budget constraints and success criteria are given.	.492	.592		
	12.	I can express evidences that I collected for a problem solution as graph or table.	.576	.585		
	13.	I can produce different solutions for solution of a problem.	.536	.580		
	14.	To understand an issue, I can consider all of the details related to that issue.	.526	.579		
	15.	I can present design problems that can be solved by applying scientific ideas on any subject.	.571	.561		
	16.	I can explain a topic by modeling it with all its components.	.478	.550		
	17.	I can understand abstract relations between objects and events by forming cause and effect relationship.	.384	.550		
	18.	I can recognize proportional relationships between quantities.	.487	.517		
	19.	I can gather scientific evidence to solve a problem.	.366	.498		
	20.	When developing a model, I can even consider the smallest details.	.372	.467		
Engineering - Technology	21.	I review the inconsistencies in my solution model and try to reveal its source.	.704		.784	
	22.	I can apply scientific principles to test a device to solve a problem.	.619		.755	
	23.	By analyzing a chart, I can explain the relationship between the quantities in the chart.	.672		.742	
	24.	I can design models for being understood of the idea that I proposed for solution.	.701		.716	
	25.	I can write statements that include clear reasons and related evidences to support a claim.	.639		.703	
	26.	I can apply scientific principles to build a device for a solution of a problem.	.559		.695	
	27.	I can logically arrange evidences of a claim.	.604		.662	
	28.	In order to understand a topic, I can collect related information in a model.	.633		.662	

	29. I can state the relationship by establishing connection between claims, reasons and evidences.	.555		.614	
	30. I can evaluate the evidence of alternative / counter-claims and, if necessary, accept it.	.574		.612	
	31. I can apply scientific principles to design a device to solve a problem.	.399		.564	
	32. I can test a model in the laboratory by setting up an experimental setup.	.430		.543	
	33. I can consider that realistic evidence is important to achieve the right solution.	.555		.539	
	34. I can suggest a mathematical model to solve a problem.	.461		.517	
	35. I can make conclusions using data from an example given about an unknown situation.	.433		.493	
Mathematics	36. I can use the properties of integers to synchronize two sides of an equation.	.661			.806
	37. I can write simple equations to solve problems.	.568			.687
	38. I am aware that positive and negative numbers can indicate a direction. (for example, -3 degrees below zero, above +4 degrees)	.436			.624
	39. I use proportioning and reasoning to solve math problems.	.532			.598
	40. I can use mathematical expressions to describe a simple model related to science fields.	.495			.592
	41. I use proportioning and reasoning to solve the problems that I encounter in daily life.	.505			.577
	42. I can work collaboratively with my other friends in a research project designed to solve a problem.	.430			.516
	43. I can use statistical interpretation methods (mean, standard deviation, etc.) so that the numerical data I obtain can be the answer to a question.	.411			.509
		Eigenvalues	9.49	8.83	4.96
		Explained variance	22.06	20.52	11.53

As it is seen in Table 2, Science factor of the scale includes 20 items and their factor loadings are ranging from 0.467 to 0.758. This factor's eigenvalue in the scale is 9.49; its portion in the total variance is 22.06%. Engineering-technology factor includes 15 items. These items factor loadings are ranging from 0.493 to 0.784. This factor's eigenvalue in the scale is 8.83; its portion in the total variance is 20.52%. Mathematics factor includes 8 items. These items factor loadings range from 0.509 to 0.806. This factor's eigenvalue in scale is 4.96; its portion in the total variance is 11.53%.

Item-Factor Total Correlations

In this part, correlation between scores of each item in factors and factor scores was calculated and level of serving for general purpose was tested for each item. Item-factor correlation values of each item are presented in Table 3.

Table 3. Item-factor scores correlation analysis

F1 Science		F2 Engineering-Technology		F3 Mathematics	
I	**r**	**I**	**r**	**I**	**r**
1	.826**	21	.825**	36	.763**
2	.790**	22	.743**	37	.746**
3	.758**	23	.805**	38	.595**
4	.657**	24	.844**	39	.733**
5	.766**	25	.789**	40	.707**
6	.732**	26	.734**	41	.699**
7	.744**	27	.773**	42	.677**
8	.745**	28	.794**	43	.653**
9	.732**	29	.755**		
10	.736**	30	.758**		
11	.698**	31	.641**		
12	.761**	32	.645**		
13	.725**	33	.733**		
14	.727**	34	.647**		
15	.743**	35	.640**		
16	.613**				
17	.697**				
18	.691**				
19	.610**				
20	.618**			N=723; **=p<. 001	

As it is seen in Table 3, item test correlation coefficients for the first factor is ranging from 0.610 to 0.826; for the second factor from 0.641 to 0.805 and for the third factor from .595 to 763. Each item has a positive and significant correlation with the total factor (p<.001). Therefore, it can comfortably be stated that each item is appropriate for the factor that it is placed.

Item Discrimination

Item discriminations of the scale items were calculated. For this purpose, first, raw scores that were gathered from the scale sorted in a descending order. Then, groups of lower and upper groups were determined with 195 individuals from lower 27% and upper 27%. Independent sample t-test scores were calculated by using total scores of groups. *t* values related to discrimination powers and significance of them are presented in Table 4.

Table 4. Item Discrimination Powers

F1 Science		F2 Engineering-Technology		F3 Mathematics	
I	**t**	**I**	**t**	**I**	**t**
1	27.959*	21	20.651*	36	7.590*
2	26.445*	22	16.762*	37	13.823*
3	26.390*	23	18.720*	38	10.57*
4	15.377*	24	21.895*	39	22.259*
5	25.469*	25	18.909*	40	13.987*
6	22.319*	26	16.957*	41	13.201*
7	21.099*	27	22.333*	42	13.471*
8	19.082*	28	21.113*	43	12.766*
9	27.841*	29	21.569*		
10	22.692*	30	20.461*		
11	20.503*	31	15.879*		
12	27.305*	32	18.217*		
13	19.856*	33	21.113*		
14	20.394*	34	15.092*		
15	25.594*	35	22.338*		
16	20.049*			**F1**	46.560*
17	21.148*			**F2**	33.568*
18	23.924*			**F3**	20.957*
19	14.746*			**FT**	44.946*
20	14.431*			***df: 388; p<.001**	

As it is seen in Table 4, independent sample t-test values related to 43 items, factors and total scores are ranging from 7.590 to 27.959. *t* value of the scale in general is 44.946. *t* values related to factor scores range from 20.946 to 46.560. Each difference is significant (p<.001). In accordance with that it can be shared that discrimination of the scale and each of its items are high.

Findings Related to Reliability of the Scale

Internal Consistency

Cronbach's alpha reliability coefficient, split-half reliability correlation, Spearman-Brown formula and Guttmann split-half reliability formula were used to calculate reliability of the scale considering the whole of the scale and its factors. Reliability analysis results considering the whole of the scale and its factors were summarized in Table 5.

Table 5. Reliability analysis results considering the whole of the scale and its factors.

Factor	Number of items	Two congruent halves correlation	Sperman Brown	Guttmann Split-Half	Cronbach's Alpha
Science	20	,887	,940	,938	,950
Engineering – Technology	15	,912	,954	,944	,940
Mathematic	8	,737	,849	,849	,848
Total	43	,852	,920	,918	,969

According to Table 5, split-half correlation of the scale is .852; Spearman-Brown reliability coefficient is .920; Guttmann Split-Half value is .918; Cronbach's alpha reliability coefficient is .969. Together with that, it is observed that split-half correlation related to factors range from .737 to .912; Spearman-Brown values range from .894 to .954; Guttmann Split-Half values range from .849 to .9382; and Cronbach's alpha reliability coefficients range from .848 to .950. Parallel to these results, it can be highlighted that both the whole of the scale and factors of it can make consistent measurements.

Stability Level

Stability level of the scale was determined by conducting test-retest method. After three weeks of the implementation, 29 students retook the 43-itemed final form of the scale. After both administration of the scale, correlation between the scores with respect to both each item and whole of the scale were measured. Thereupon, both each item's and the whole of the scale's ability to make stable measurements were tested and results are summarized in Table 6.

Table 6 Test-retest results of the items of the scale.

F1 Science		**F2 Engineering-Technology**		**F3 Mathematics**	
I	**r**	**I**	**r**	**I**	**r**
1	.849**	21	.531**	36	.349*
2	.523**	22	.404*	37	.640**
3	.889**	23	.340*	38	.393*
4	.686**	24	.639**	39	.506**
5	.572**	25	.342*	40	.477*
6	.618**	26	.524**	41	.684**
7	.900**	27	.661**	42	.895**
8	.651**	28	.526**	43	.400*
9	.364*	29	.386**		
10	.457*	30	.874**		
11	.887**	31	.465*		
12	.321*	32	.536**		
13	.398*	33	.471**		
14	.351*	34	.777*		
15	.888**	35	.443*		
16	.864**			**F1**	.833**
17	.740**			**F2**	.627**
18	.577**			**F3**	.741**
19	.549**			**FT**	.812**
20	.536**			**N= 29; *=p<0.05 **=p<0.001**	

As presented in Table 6, correlation coefficients of each item of scale ranges from .321 to .900, which were produced by test-retest method, and each correlation was significant and positive. Correlation coefficients of scale's factors, which were obtained by the same method, range from .741 to .833. It is observed that correlation related to the total score is .892 and each correlation was significant and positive. Whence it can be declared that this scale makes stable measurements.

4. Result, Discussion and Suggestions

In this study, "Basic STEM Skill Levels Perception Scale" was developed to identify students' basic STEM skill levels according to their perception. This scale is a 7-point likert-type scale consisted of 43 items under three factors. Each item under the factors was scaled between "(1) strongly disagree" to "(7) strongly agree". To identify the factor structure of the scale exploratory factor analysis was conducted. According to this exploratory factor analysis, the factor loadings of items, eigenvalues of factors and explained variances of this scale it can be said that it has construct validity. Item factor correlations of each items of the scale were calculated to identify at what level these items were able to measure the skills of the factors they belonged to. Finding correlation between scores gained from each item and the score of the factors that the item belongs to was used as a criterion to identify the level of service of each item to general purpose of each factor (Balcı, 2009). Accordingly, correlation between scores gained from each item and the score of factors that the item belongs to range from 0.595 to 0.826. Thus, it can be said that each item and the factor of the scale significantly serve for measuring skills aimed to evaluate scale-wide and item discrimination level of each item is appropriate. Internal consistency coefficient of the scale was calculated by Cronbach Alpha formula and was determined as .969. Hence, it is identified that both each factor and scale able to make consistent measurements. According to Murphy and Davidshoper (1988) Cronbach Alpha value which is higher than .90 indicates high level reliability. Stability of the scale was identified by conducting test-retest method and results showed that, between these two applications, correlation coefficients of factors ranged .320 to .900 and correlation related to scores was .812. As a result, it can be said that this scale is able to make reliable measurements. In addition, it can be said that "Basic STEM Skill Levels Perception Scale" is a valid and reliable scale to measure university students' basic STEM skills according to their self-perception. Considering the general structure of STEM approach, it is thought that using factors of the scale separately would not be appropriate. In this line of thought it is recommended to use this scale to measure university students' perceived skills.

References

Balcı, A. (2009). *Sosyal bilimlerde araştırma: Yöntem, teknik ve ilkeler* [Research in social science: Methods, techniques and principles]. Ankara: PegemA Pub.

Brown, W. (2015). *Introduction to Algorithmic Thinking*. Available at: https://www.coursehero.com/file/31890645/Introduction-to-Algorithmic-Thinkingdocx/

Büyüköztürk, Ş. (2002). *Sosyal bilimler için veri analizi el kitabı* [Data analysis handbook for social sciences]. Ankara: PegemA Press.

Bybee, R. W. (2013). *The case for STEM education: challenges and opportunities*. Virginia: NSTA Press

Carmines, E.G. & Zeller, R.A. (1982). *Reliability and validity assessment. 5th ed.* Beverly Hills: Sage Publications Inc.

Çepni, S. (2018). *Kuramdan Uygulamaya STEM Eğitimi [From theory to practice STEM education]*. Ankara, PegemA.

Eroğlu, A. (2008). Faktör analizi [Factor analyses]. In: Kalaycı, Ş. (Ed.), *Spss uygulamalı çok değişkenli istatistik teknikleri* [Statistics Techniques with Multi Variable in SPSS Applications] (pp. 321-331). Ankara: Asil Publishers.

Field, A. (2000). *Discovering Statistics using SPSS for Windows*. London: Thousand Oaks- New Delhi: Sage Pub.

Gelen, İ. (2017). P21- 21st Century Skill Frameworks in Curriculum and Instruction (USA Practices). *Journal of Interdisciplinary Educational Research, 1*(2), 15-29.

Gökkurt, B., Örnek, T., Hayat, F. & Soylu, Y. (2015). Assessing Students' Problem-Solving and Problem-Posing Skills. Bartın University Jounal of Faculty of Education, 4(2), 751-774. doi:10.14686/buefad.v4i2.5000145637

Gorsuch, R. L. (1983). Factor analysis. Hillsdale: Lawrence Erlbaum Associates.

Gülhan, F. & Şahin, F. (2016). The effects of science-technology-engineering-math (STEM) integration on 5th grade students' perceptions and attitudes towards these areas. *International Journal of Human Sciences, 13*(1):602- 620. doi:10.14687/ijhs.v13i1.3447

Günüç, S., Odabaşı, H. F.& Kuzu, A. (2013). The Defining Characteristics of Student of The 21. Century By Student Teachers: A Twitter Activity. *Journal of Theory and Practice in Education, 9*(4), 436-455.

Hernandez, P. R., Bodin R., Elliott, J. W., Ibrahim B., Rambo-Hernandez, K. E., Chen T. W. & Miranda M. A. (2014). Connecting the STEM dots: measuring the effect of an integrated engineering design intervention. *International Journal Technology Design Education*. 24, 107-120, doi:10.1007/s10798-013-9241-0

Honey, M., Pearson, G. & Schweingruber, H. (Eds). National Academy of Engineering and National Research Council (2014). STEM integration in K12 education: Status, prospects, and an agenda for research. Washington D. C. : The National Academies Press.

Keçeci, G., Alan, B. & Kırbağ Zengin, F. (2017). STEM Education Practices with 5th Grade Students. *Ahi Evran University Journal of Kırşehir Educational Faculty (KEFAD), 18*, 1-17.

Kline, P. (1994). *An easy guide to factor analysis*. London and New York: Routledge.

MNE (2011). *21. Yüzyıl Öğrenci Profili* [21. Century Student Profile]. Ankara: Turkey Ministry of National Education

Murphy, K. & Davidshofer, C. (1988). *Psychological testing: Principles and applications*. Englewood Cliffs, NJ: Prentice-Hall.

NGSS. (2018). DCI Arrangements of the Next Generation Science Standards. Available at: https://www.nextgenscience.org/sites/default/files/NGSS%20DCI%20Combined% 2011.6.13.pdf

Rogers, C. & Portsmore, M. (2004). Bringing engineering to elementary school. *Journal of STEM Education, 5*(3), 17-28.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis. *Personality and Social Psychology Bulletin*, 28, 1629–1646. doi:10.1177/014616702237645

Scherer, R.F., Wiebe F.A., Luther, D. C. & Adams J. S. (1988). Dimensionality of coping: Factor stability using the ways of coping questionnaire. *Psychological Reports* 62(3), 763-770. PubMed PMID: 3406294. doi:10.2466/pr0.1988.62.3.763

SNAP (2018). SNAP Assessment for NGSS. Available at: https://snapgse.stanford.edu/snap-assessments-ngss

Stinson, K., Harkness, S., Meyer, H. & Stallworth, J. (2009). Mathematics and science integration: models and characterizations. *School Science and Mathematics*, 109(3), 153-161. doi:10.1111/j.1949-8594.2009.tb17951.x

Sullivan, F. R. (2008). Robotics and science literacy: Thinking skills, science process skills and systems understanding. *Journal of Research in Science Teaching*, 45(3), 373–394. doi:10.1002/tea.20238

Tatlıdil, H. (2002). *Uygulamalı çok değişkenli istatistiksel analiz* [Applied multivariate statistical analysis]. Ankara: Akademi Pub.

Whittle, S.R., Pell, G. & MurdochEaton, D.G. (2010). Recent changes to students' perceptions of their key skills on entry to higher education. *Journal of Further and Higher Education*, 34(4), 557-570. doi:10.1080/0309877X.2010.512082

Yakman, G, (2008). STEM Education: An overview of creating a model of integrative education. Pupils Attitudes Towards Technology Annual Proceedings. Netherlands.

Yıldırım, B. & Selvi, M. (2017). An experimental research on effects of stem applications and mastery learning. *Journal of Theory and Practice in Education*, 13(2), 183-210. doi:10.17244/eku.310143

Yılmaz, A., Gülgün, C. & Çağlar, A. (2017). Teaching with STEM Applications for 7th Class Students Unit of "Force and Energy": Let's Make a Parachute, Water Jet, Catapult, Intelligent Curtain and Hydraulic Work Machine (Bucket Machine) Activities. *Journal of Current Researches on Educational Studies, 7*(1), 97-116.