

STEM Club Evaluation Scale: Validity and Reliability Study

Hasan Gokce*¹, Seyide Eroglu¹, Melek Karaca², Oktay Bektas³

¹Ministry of National Education, Kayseri, Turkey

²Institute of Educational Sciences, Erciyes University, Kayseri, Turkey

³Faculty of Education, Erciyes University, Kayseri, Turkey

*Corresponding author: hasangokce3838@gmail.com

ABSTRACT In STEMNET's report, 76% of 500 teachers interviewed stated that joining the STEM Club increased students' ability to solve real-world problems. This study aims to develop a valid and reliable measurement tool for evaluating STEM clubs. The research sample consisting of 149 teachers who carry out STEM club activities in schools in Turkey was determined using the purposive sampling method. Content and construct validity and reliability analyses have been performed for this purpose. To ensure content validity, (1) a pool of questions based on the literature was created, (2) draft scale items were determined, (3) an expert was allowed to check them, and (4) item difficulty and discrimination index were calculated. To ensure construct validity, (1) exploratory factor analyses (EFA) and (2) confirmatory factor analyses (CFA) were performed on both the same and different samples. As a result of the analyses, having the same data set be analyzed with different software was sufficient for verifying the factor structure. A three-factor structure consisting of 29 items was obtained, which explains 52% of the variance. Cronbach's alpha of reliability for the overall scale was calculated as .92. As a result, a valid and reliable scale was determined to have been developed for researchers and program practitioners to evaluate STEM clubs. Suggestions have been made that the scale can be used on STEM clubs at the provincial, district, and school levels to determine their efficiency and productivity.

Keywords Science Education, STEM, Scale Development, Validity, Reliability

1. INTRODUCTION

STEM education emerged as an educational reform. Turkey adopted it in 2014 and entered it onto the national education policy agenda. Therefore, the number of studies on STEM education and its importance has also increased (Corlu, Capraro & Capraro, 2014; Ministry of National Education [MoNE], 2016; Turkish Industry & Business Association [TUSIAD], 2014). This issue became official and started to be discussed in education circles for the first time with the STEM Education Report published by the MoNE-affiliated General Directorate of Innovation and Educational Technologies (2016). This report emphasized that STEM education urgently needs to be included in Turkey's current education system for sustainability in the economy; an action plan was created, and solution suggestions were listed. Some of the solution suggestions included in the action plan involve: establishing and disseminating STEM Centers in provinces, encouraging educational researchers to conduct research in this field, supporting teachers with pre-service and in-service training, updating curricula, and providing an environment

for students to perform STEM education activities regardless of time or place (MoNE, 2016).

Considering the STEM Education Report, studies have been initiated to integrate STEM education into lessons. First of all, changes were made to the curricula. Accordingly, the objectives and achievements of the science course curriculum were revised in 2018 and harmonized with STEM education. Thus, the aim is also to create a suitable framework for integrating STEM into lessons. According to the 2018 curriculum, students are given opportunities to propose solutions to daily life problems, experience engineering applications, and use different disciplines in their courses (MoNE, 2018).

However, when examining the studies on teachers' or teacher candidates' opinions on how to conduct STEM activities in classes, they are seen to state the content and duration of the course to be a major obstacle in making the applications (Siew, Amir, Chong, 2015). When examining

Received: 28 October 2021

Revised: 16 March 2022

Published: 27 July 2022

the studies on the problems experienced in STEM education in Turkey (Akgunduz et al., 2015; Altunel, 2018; Eroglu & Bektas, 2016), teachers and teacher candidates are seen to stated being unable to carry out studies on STEM education due to the course load and curriculum density. This problem can be solved by changing the curriculum, course content, and class hours. However, because this solution implies long-term radical change, the most practical way to apply STEM is to conduct activities outside the classroom. For this reason, STEM studies mainly use out-of-school learning environments (Baran, Bilici, Mesutoglu & Ocak, 2016; Kalkan & Eroglu, 2017; National Research Council, 2015). Out-of-school learning environments have significant potential for increasing student learning and providing them with a rich learning environment (Robelen, 2011). Through the activities carried out in after-school programs, students acquire various skills, produce many solutions to daily life problems, and learn to cooperate and communicate (Mahoney, Parente & Lord 2007).

Out-of-school learning environments cover a wide area, such as social, cultural, and technical trips around the school, field studies, project studies, sports activities, nature training, or club activities (Karadogan, 2016). By applying STEM activities in out-of-school learning environments, students are supported in terms of career choices, meaningful learning, and interest in science lessons (Dabney et al., 2012). The related literature states that STEM activities in out-of-school learning environments direct students' career plans to STEM fields (Dabney et al., 2012). In addition, it is underlined that STEM activities carried out in out-of-school learning environments are essential in providing deep learning for students (Bybee, 2001).

One of the out-of-school learning environments where STEM activities are carried out is social clubs (Afterschool Alliance, 2015; Bell, Lewenstein, Shouse & Feder, 2009). STEM clubs are expressed as flexible working environments created without regard to time or place where out-of-school STEM studies are carried out (Blanchard, Hoyle & Gutierrez, 2017). While STEM clubs develop students' emotional skills such as belonging and peer-to-peer communication, they also enhance students' 21st-century skills and help them learn current content. In addition, the activities carried out in STEM clubs support the formation of career awareness in students and their orientation toward STEM professions (Blanchard, Hoyle & Gutierrez, 2017).

In the related literature, many studies reveal the positive effects of STEM studies carried out under social clubs on students (Ayers, Wade-Jaimes, Wang, Pennella & Pounds, 2020; Baldrige, Nutt, Vaughn, Hartley-Lewis & Amos, 2019; Lipuma, Bukiet & Leon, 2021). As one of these studies, the STEM club study conducted by Ayers, Wade-Jaimes, Wang, Pennella & Pounds (2020) with different

partner schools enabled students to face real-life problems. They called St. Jude STEM Club (SJSC), and students conducted pediatric cancer research with accurate data for a 10-week. As a result of the study determined that students' attitudes toward science changed positively, their interest in the STEM field, and their participation in club activities increased (Ayers, Wade-Jaimes, Wang, Pennella & Pounds, 2020). In another study, activities in cooperation with a university and a high school within the STEP (The Student and Teacher Enhancement Partnership) program conducted by Baldrige, Nutt, Vaughn, Hartley-Lewis & Amos (2019). As a result of the activities carried out with mentors from the university, they determined that many skills, including scientific literacy, writing, and continuing science skills, developed.

Based on those above, it can be stated that while STEM activities carried out under social club activities eliminate the time problem, it also supports formal education and contributes to the development of students. For this reason, STEM clubs are seen as an appropriate way to carry out STEM studies effectively (Straw, Branson, Neumann & Dickinson, 2011). Based on this, MoNE YEGITEK sent an official letter dated May 2018 to all schools in Turkey. Accordingly, the official article stated that STEM clubs might be formed to carry out STEM studies more effectively in schools. In this direction, schools have started to form STEM clubs as of the 2018-2019 academic year (Coskun, Alakurt, & Yilmaz, 2020).

STEM clubs have been formed based on MoNE's (2017) Educational Institutions Social Activities Regulation. When examining this regulation, it is seen to involve general principles. Although these are the same regarding the basic principles of STEM club activities in schools, no legislation exists that sets out STEM clubs' working structures or activity frameworks. That suggests that differences may exist in how content is applied. Determining the standards for the studies carried out in STEM clubs and establishing STEM clubs within the framework of scientific and other criteria are extremely important for carrying out school applications without interruption (Vural, 2018).

When examining the relevant studies in the literature, out-of-school studies are found to have been conducted with STEM clubs in the title (Ferrara et al., 2017; Gottfried & Williams, 2013; Sahin, 2013). However, no studies were found to have examined, evaluated, or determined the deficiencies of STEM club studies. For this reason, the current situation must be revealed in all its aspects to organize STEM club activities and eliminate their problems. Obtaining the opinions of the teachers who carry out STEM club work on the subject is essential for determining the current situation. While the effects of STEM club activities on students are frequently investigated in the related literature (Ferrara et al., 2017; Gabrielson, Strachan, Warner & LaFleche, 2009), no study

is found to have obtained the opinions of teachers who carry out STEM club activities. However, we think it is more appropriate to start researching STEM clubs from teachers since they are the practitioners of the programs who work in the field and can best identify the positive and negative aspects of the studies. In addition, we believe that it would be more appropriate to start the application with the students, starting with the teachers, since they are the teachers who know and observe their students best. For this reason, primarily teachers were studied; In addition to the teacher element, detailed analyses were made in the study in terms of planning and implementation and student dimensions, which are other important elements.

In the related literature, there are many scale development studies in the field of STEM (Cevik, 2017; Faber et al., 2013; Guzey, Harwell & Moore, 2014; Milner, Horan & Tracey, 2014). When these studies are examined in detail; we have seen that they focus mostly on cognitive and affective areas such as awareness of STEM fields (Cevik, 2017), attitude towards STEM (Buyruk & Korkmaz, 2016, Faber et al., 2013), self-efficacy (Milner, Horan & Tracey, 2014). However, to increase the capacity and impact of STEM studies, the studies should be evaluated according to specific criteria. The current situation should be revealed in all aspects to identify the good or bad aspects of STEM studies in or out of school and make improvements if necessary. In this context, evaluation tools for STEM studies are needed, but instead

of special scales, alternative assessment and evaluation tools are preferred (Zengin, Kaya & Pektaş, 2020). However, it is important to determine the effectiveness of different STEM practices, and situation-specific scales need to be developed. Still, no measurement tool can evaluate all aspects of the STEM club activities that teachers carry out in their classrooms.

Therefore, the decision has been made to conduct such a study for the following reasons: to be able to carry out investigations on STEM education through out-of-school clubs, STEM clubs are a way to adapt learning environments to STEM education, a limited number of studies are found in the literature on STEM clubs, legislation that sets out the operational structures and activity frameworks of STEM clubs is lacking and creates non-standard situations in practice, no studies are found on measuring the effectiveness of STEM clubs, and no measurement tool exists that can be used to evaluate STEM club activities.

This research aims to develop a measurement tool that will reveal the level to which the studies carried out in the STEM clubs established in their schools have been implemented, the problems experienced in the implementation process, and the advantages of the applications. Validity and reliability studies of the developed measurement tool have been carried out for this purpose. In other words, this research aims to bring a practical and easily applicable measurement tool to the

Table 1 Draft scale in blueprint format

Dimension	No	Items	1	2	3	4	5
Teacher (Factor 1)	1.	Club activities are carried out regularly in my school.					
	2.	Teacher selection for the STEM club in my school is done voluntarily .					
	3.	I have sufficient knowledge about STEM education.					
	4.	Student selection for the STEM club at my school is not voluntary.					
	5.	I do not have any difficulties while carrying out STEM club activities at my school.					
	6.	I act according to the club plan while performing STEM club activities.					
	7.	The STEM club plans I use are associated with science achievements.					
	9.	The STEM club plans I use are related to information technology gains.					
	11.	STEM club activities in my school contribute to the success of the students.					
	17.	It contributes to the use of STEM club activities in my school and different teaching strategies in my lessons.					
	18.	STEM club activities in my school contribute to the students' ability to solve their daily life problems.					
	19.	It gives students an interdisciplinary perspective on STEM club work at my school.					
	21.	Students take an active role in STEM club activities at my school.					
	22.	STEM club activities in my school contribute to the development of positive attitudes of students towards school.					
	24.	STEM club activities in my school contribute to the career choice of students.					
	27.	I would like to take an active part in the STEM club every year.					
	30.	When organizing STEM club work, I create the club plan myself.					
32.	The STEM club plans I use are associated with technology design achievements.						

Table 1 Draft scale in blueprint format (*Continued*)

Dimension	No	Items	1	2	3	4	5
Planning and implementation (Factor 2)	10.	My school's facilities are not sufficient to run STEM club activities.					
	12.	I can easily obtain materials used in STEM club studies.					
	14.	Different types of STEM club activities are not carried out in my school.					
	23.	The time allocated for STEM club activities carried out in my school is not enough.					
	29.	The fact that STEM club activities are seen as a lesson activity and carried out in one lesson prevents the effectiveness of the activities.					
	31.	During STEM club activities, most of the time, there are no activities and the studies remain on paper.					
	33.	During STEM Club activities, there is no cooperation with official and voluntary organizations in the district.					
Student (Factor 3)	8.	STEM club activities in my school do not contribute to the students' ability to use technology effectively.					
	13.	STEM club activities at my school have no contribution to students' discovery of their talents.					
	15.	Few students participate in STEM club activities.					
	16.	The STEM club plans I use are not suitable for the student level.					
	20.	STEM club activities in my school do not contribute to the development of different materials in my lessons.					
	25.	I do not think that STEM club activities at my school improved my communication skills with my students.					
	26.	The STEM club activities at my school do not contribute to the students' ability to think like scientists.					
	28.	Students who carry out STEM club activities at my school are expected to meet certain criteria.					
	34.	The STEM club plans I use are not associated with math achievements.					

literature to enlighten authorities on how to create a special framework plan for STEM clubs. Furthermore, to determine the functioning of social clubs in schools, enable administrators to determine the effectiveness of club activities, and reveal practitioners positive and negative experiences. Furthermore, the study will help bring new scales to the field, determine the strengths and weaknesses of some special applications, make the necessary revisions, carry out studies without interruption, and increase the maximum benefit of the applications for students. Finally, the study will also facilitate the work of experts in this field.

The research questions determined in line with the aims of the study are as follows;

1. Is the scale developed for determining the level of teachers' implementation of STEM club activities valid?
2. Is the scale developed to determine the level of teachers' implementation of STEM club activities reliable?

2. METHOD

This section explains the research design, universe, sample, data collection, and analysis.

2.1. Research Design

This study has chosen the survey design, a quantitative research method. Survey designs are generally defined as

the numerical expression of attitudes, tendencies, and opinions about the community, using the answer options determined by the researcher from a community (Creswell, 2017; Fraenkel, Wallen & Hyun, 2012). The survey design has been preferred within the scope of the current research to numerically express the STEM Club Evaluation Scale (SCES) scores the teachers who constitute the study sample received and to perform analyses with these scores. The survey design meets these needs.

2.2. Population and Sample

This research has identified the accessible population as the teachers who carry out STEM club activities in schools in Turkey. To make generalizations in validity and reliability studies, it is necessary to reach five times the number of items (Tabachnick & Fidell, 2007). For the results to be statistically significant for SCES, a sample of five times the 34 items in the scale was tried to be reached, and this number was determined as 149 (Hair, Black, Tatham & Anderson, 2019). Official registration of the number of teachers in the accessible population is not permitted. For this reason, the authors could not study with at least 10% of the population. The study has preferred purposive sampling. Purposeful sampling is determining a sample of people and situations suitable for the research (Johnson & Christensen, 2014). Purposive sampling is preferred because the investigation is conducted with

teachers who carry out STEM club activities in the schools where they teach.

2.3. Data Collection Tool

The scope of the current research aims to develop a data collection tool. The authors have conducted validity and reliability studies of the SCES as a measurement tool. The draft scale created by the authors based on science education and STEM literature is given in Table 1 in blueprint format. The authors wanted to develop a measurement tool to evaluate the effectiveness of out-of-school STEM practices. Therefore, it was paid attention to creating items for the teachers who are the implementers of the program and the student who is the addressee of the program, which is suitable research. Since the efficiency of STEM clubs requires good planning and implementation, articles for planning and implementation are also written. In short, it was predicted that the draft scale consisted of teacher, student, planning, and implementation dimensions.

2.4. Data Collection

At the level that teachers implement STEM club activities, the following steps have been followed within the scope of the SCES development studies:

1. A data collection tool has been prepared as a result of the relevant literature review.
2. The sample over which the study will be conducted has been determined.
3. The authors transferred the SCES to the Google Questionnaire application and allowed it to be filled out online.
4. After the teachers filled out the SCES online, the authors transferred the data obtained from the Google Survey application to the package program SPSS 25.

2.5. Data Analysis

The obtained data have been analyzed using the package programs SPSS.25 and LISREL 8.80. The study data have been evaluated at a significance level of $p < .05$. It is explained in detail in the Findings section. General headings are given in the data analysis to avoid repetition. To provide evidence for the validity and reliability of the scale, DeVellis' (2014) eight steps for scale development were adhered to, and the following analyses have been made in order:

- Literature review, question pool creation, expert opinion, and item index analyses to ensure content validity;
- Explanatory and confirmatory factor analyses over the same and different samples to construct validity;
- Cronbach alpha reliability analysis was performed to ensure reliability.

It is common in the literature that confirmatory factor analysis should be done with data obtained from a different sample than the sample in which exploratory factor analysis was performed. This study discusses the accuracy of this common idea by using the same and different data sets for

exploratory and confirmatory factor analysis. The expression "same sample" means that the data set used in exploratory factor analysis is also used in confirmatory factor analysis. The expression "different sample" means that the data set used in exploratory factor analysis is different from that used in confirmatory factor analysis.

The draft SCES is a 5-point Likert-type scale consisting of 34 items. The scale has 21 positive items (Items 1, 2, 3, 5, 6, 7, 9, 11, 12, 15, 17, 18, 19, 21, 22, 24, 25, 27, 28, 30, 31, and 32) and 13 negative items (Items 4, 8, 10, 13, 14, 16, 20, 23, 25, 26, 29, 33, and 34). In scale development studies, one way to check that participants read and answer the scale items is to use positive and negative items that measure the same dimension. Scales containing positive and negative statements are widely used to lessen the acquiescent response bias (Qasem & Gul, 2014). Before the analysis, negative items were reverse coded using the recode command. The five points of "strongly disagree" (1 pts.), "disagree" (2 pts.), "undecided" (3 pts.), "agree" (4 pts.), and "strongly agree" (5 pts.) have been used to determine the level to which each item in the data collection tool has been realized.

The Kolmogorov-Smirnov and Shapiro-Wilk tests have been used to determine the normal distribution assumption of the SCES scores from the research data. Histograms and skewness-kurtosis coefficients, mean, mode, median, and standard deviation values have been examined. For teachers' SCES scores to meet the assumption of normal distribution, the mean and median values should have similar values, and the skewness-kurtosis coefficients should be between -2 and +2 (George & Mallery, 2016). Based on the sample size, the decision was made to use either the Kolmogorov-Smirnov or Shapiro-Wilk test. This study used the Kolmogorov-Smirnov test because the sample size was greater than 50 (Buyukuzturk, Kilic-Cakmak, Akgun, Karadeniz & Demirel, 2016). In data analysis, the data are assumed to have normal distribution when $p > 0.05$ (Pallant, 2016)

3. FINDINGS

3.1. Normality Analysis Findings

The mean, median, and mode values for each item in the developed scale were close, and the kurtosis and skewness values are between -2 and +2. Therefore, because the items on the draft scale have a normal distribution, no items were removed (George & Mallery, 2016). The Kolmogorov-Smirnov test results were also in the desired range (Pallant, 2016). In addition, the entire sample within the scope of construct validity was determined to have a normal distribution for the EFA and CFA. To perform CFA on a different sample, the sample was split in two to ensure normal distribution. The average distribution of the samples was determined by examining the histograms to show the mean, median, and mode values to be close to

one another and the kurtosis-skewness values to vary between -2 and +2 (Fraenkel & Wallen, 1996).

3.2. Validity Analysis Findings

Validity is the degree to which the measurement tool serves its purpose (Turgut & Baykul, 2015). Content and construct validity analyses were made to ensure the validity of the developed scale. The obtained results are presented in order.

3.2.1. Content Validity Findings

The items on the SCES scale were created based on constructivist theory, the theoretical framework of which is the philosophical approach on which STEM education is based. In addition, DeVellis' (2014) 8-step scale development method has been followed (see Figure 1).

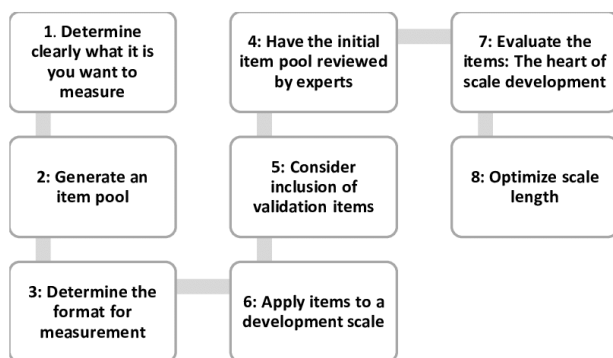


Figure 1 Scale development steps

Content validity can be defined as the extent to which the items making up the test represent the behavioral universe to be measured. Therefore, the content and framework must be consistent for a study to have content validity (Fraenkel & Wallen, 1996). In this context, how the

SCES items were created is explained in detail. In other words, while preparing the scale, the structure to be measured was defined, an item pool was created, which type of scale it would be was decided, expert opinions were sought, and the items were revised and finalized (DeVellis, 2014).

In creating the scale, the literature on the subject was first reviewed (Ferrara et al., 2017; Gonsalves, Rahm, & Carvalho, 2013; Gottfried & Williams, 2013; Sahin, Ayar, & Adigüzel 2014). Based on the literature, an interview form consisting of 22 open-ended questions and various probes was created at the beginning of the study. Next, the form was revised, reduced to 17 items, and presented to an expert for their opinions. Finally, the interview form was examined by a science education specialist and an assessment and evaluation specialist. As a feedback result received from the expert and the change of opinion to apply the study to a wider audience, the decision was made to convert the interview form to a Likert-type scale.

In the beginning, 28 statements had been written based on the literature. Before receiving its final form, opinions from three experts (an academician in science education, assessment and evaluation, and a science teacher) were obtained regarding the scale. Line data obtained from the experts' opinions, the scale was re-examined by the researchers in terms of clarity, appropriateness, and adequacy of the questions. Some items were changed, others removed, and still a few others added in line with the experts' opinions to arrive at a 34-item scale. The item "I create the club plan myself when organizing STEM club activities" was found to be excessive and removed from the scale. Apart from this, other expressions were decided to be reverse coded. For example, "STEM club activities in

Table 2 Item difficulty and discrimination index values for the scale

Item Number	Item Discrimination Index	Item Difficulty Index	Item Number	Item Discrimination Index	Item Difficulty Index
1	.76	.54	18	.50	.75
2	.58	.68	19	.34	.83
3	.50	.72	20	.32	.84
4	.37	.45	21	.63	.68
5	.76	.57	22	.53	.74
6	.58	.68	23	.61	.49
7	.50	.64	24	.55	.67
8	.21	.84	25	.32	.71
9	.53	.66	26	.34	.83
10	.66	.41	27	.45	.78
11	.66	.67	28	.07	.33
12	.79	.53	29	.50	.38
13	.42	.79	30	.66	.64
14	.42	.45	31	.02	.14
15	-.05	.23	32	.39	.75
16	.32	.84	33	.45	.51
17	.34	.77	34	.39	.77

my school contribute to students' ability to think like scientists" was edited to say "do not contribute." Lastly, Items 8, 10, 15, 29, 31, and 33 were added to the scale.

Some of the statements were formed by considering the literature. For example, based on the statement "Only willing students should be recruited for club work," the reverse-coded statement "Student selection for the STEM club in my school is not done voluntarily" was created (Polat, 2017). Another example statement, "You work in cooperation with public and private non-governmental organizations as well as parents" (Birturk, 2015), which is used as the new scale item "There is no collaboration with official or volunteer organizations in the district during STEM Club activities."

The constructivist theory was considered while creating the scale items as the basic philosophical approach on which STEM education is based. Accordingly, the items include interdisciplinary expressions by the nature of STEM education where the student is placed in the center to determine whether students take an active role in the process or not and their status regarding being able to use different skills and competencies. This scale is scored as a 5-point Likert-type scale ranging from strongly disagree (1) to agree (5) strongly.

3.2.2. Item Index Analysis Findings

Item difficulty and discrimination indexes have been calculated to contribute to the content validity of the developed scale (see Table 2). The criteria determined by Ebel & Frisbie (1991) were taken into consideration when evaluating the item discrimination index (see Table 3)

Table 3 Item discrimination index criterion

Item Discrimination Index Value	Evaluation
0.19 or smaller	It should never be scaled or completely corrected.
Between 0.20 and 0.29	It can be corrected and taken to the test.
Between 0.30 - 0.39	Can be tested without correction.
0.40 or higher	Very good conditions can be taken as the test.

When examining Table 3, the discrimination index for Items 8, 28, and 31 have been determined to be low ($r_8 = 0.21$, $r_{28} = 0.07$, $r_{31} = 0.02$) and the discrimination index for Item 15 to be low and negative (-0.05). After examining the other analysis results, these items were decided to be removed from the scale. The discrimination indices for other items apart from these four items vary between 0.32 and 0.79, and the item difficulty indexes vary between 0.38 and 0.84.

3.3. Construct Validity Findings

After completing the content validity, EFA and CFA were performed on the same and different samples to ensure the scale's construct validity. A scale development study uses factor analyses to determine the measurement tool's factor structure and verify a specific factor structure

(Secer, 2017). Before performing the factor analysis, Cronbach's alphas of reliability were examined for each item on the draft scale and their impact on the reliability of the entire scale if the item were to be deleted. As a result of the analyses, the reliability coefficients for Items 15, 28, and 31 on the draft scale were negative. Therefore, the decision was made not to include Item 4 ($\alpha = .069$) in the factor analysis as the value is less than .20. After removing the four items from the item index and reliability analyses, the factor analyses were performed on the remaining 30 items from the scale.

3.3.1. Factor Analysis Findings for the Same Sample

Within the scope of the current study, factor analyses have been performed on the same and different samples. This section explains the EFA and CFA performed on the same sample using the entire research sample ($N = 139$). KMO and Bartlett's test results have been calculated using SPSS 25 to check the suitability of the research data for factor analysis (see Table 4).

Table 4 KMO and Bartlett's test results for the same sample

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.890
Bartlett's Test of Sphericity	Approx. Chi-Square	2228.706
	df	406
	Sig.	.000

When examining Table 4, the KMO value has been determined to be greater than the minimum value of .60 required for analysis and to be statistically significant (Tabachnick & Fidell, 2007). Therefore, factor analysis was performed without factor limitation, and a 6-factor structure was obtained. However, the factor analysis was repeated by limiting the number of factors to three, as suggested by the scree plot, due to a large number of overlapping items and the presence of only one item in the last factor (Figure 2).

When examining Table 5, the developed scale has a structure consisting of three factors and 29 items that explain 52.02% of the total variance. When examining the

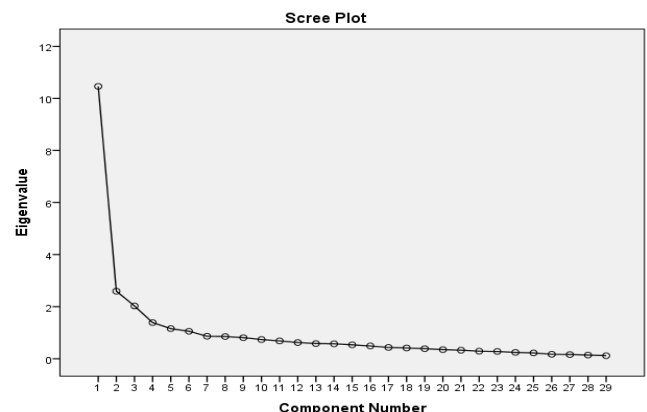


Figure 2 Scree plot for the same sample

Table 5 Total explained variance for the sces for the same sample

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.463	36.079	36.079	10.463	36.079	36.079
2	2.597	8.955	45.034	2.597	8.955	45.034
3	2.027	6.988	52.022	2.027	6.988	52.022
4	1.389	4.791	56.813			
5	1.159	3.995	60.808			
6	1.056	3.643	64.450			
7	.865	2.984	67.434			
8	.854	2.945	70.379			
9	.811	2.797	73.175			
10	.739	2.548	75.723			
11	.685	2.362	78.086			
12	.624	2.150	80.236			
13	.584	2.014	82.249			
14	.575	1.982	84.231			
15	.536	1.849	86.081			
16	.492	1.695	87.776			
17	.434	1.497	89.273			
18	.414	1.427	90.700			
19	.388	1.339	92.040			
20	.350	1.206	93.246			
21	.328	1.131	94.377			
22	.291	1.005	95.382			
23	.280	.964	96.346			
24	.245	.843	97.190			
25	.224	.772	97.961			
26	.173	.595	98.556			
27	.161	.555	99.112			
28	.140	.483	99.595			
29	.117	.405	100.000			

items contained in the factors (Table 6), Factor 1 became “teacher,” Factor 2 became “planning and implementation,” and Factor 3 became “student.” To verify the obtained structure, CFA was performed on the same sample using LISREL 8.80. The number of samples required to perform CFA was determined to be 75.75, and the analysis was performed on the people in the current study ($N = 139$). The raw data and scale items were grouped according to the factors specified in the EFA results, and syntax commands were written and made suitable for CFA. Figure 3 provides the t values obtained as a result of the CFA. Jöreskog and Sörbom (1993) drew attention to looking for the presence of a red arrow while examining these t values. When a red arrow indicates no item, the model interpretation may proceed. As can be seen in Figure 3, the analysis continued with the same items as the t values had no issues.

Factor 1 here represents the teacher, Factor 2 represents planning and implementation, and Factor 3 represents the students. The next step makes sure that each item has a factor loading value of at least .30. When

examining Figure 4, the factor loading values for all items are noted to be .30 or greater.

Factor loadings, t values, the desired criteria, and the model fit indices have been examined. First, the χ^2 value compatibility index and the ratio of this value to the degrees of freedom (df) are examined. When examining the path diagram, the χ^2 value is understood to be 790.16 ($df = 377$, $p = 0.00$, $\chi^2 / df = 2.09$). The χ^2 value is understood to be low but significant in terms of the model fit indices; χ^2 / df is less than three, and the current scale has a perfect fit (Jöreskog & Sörbom, 1993). Although the initial data on the model fit was good, other fit indices have also been examined as χ^2 has a significant value (see Table 7). The SCES was determined during its development phase to consist of 29 questions and three factors resulting from the EFA; the CFA has confirmed this structure.

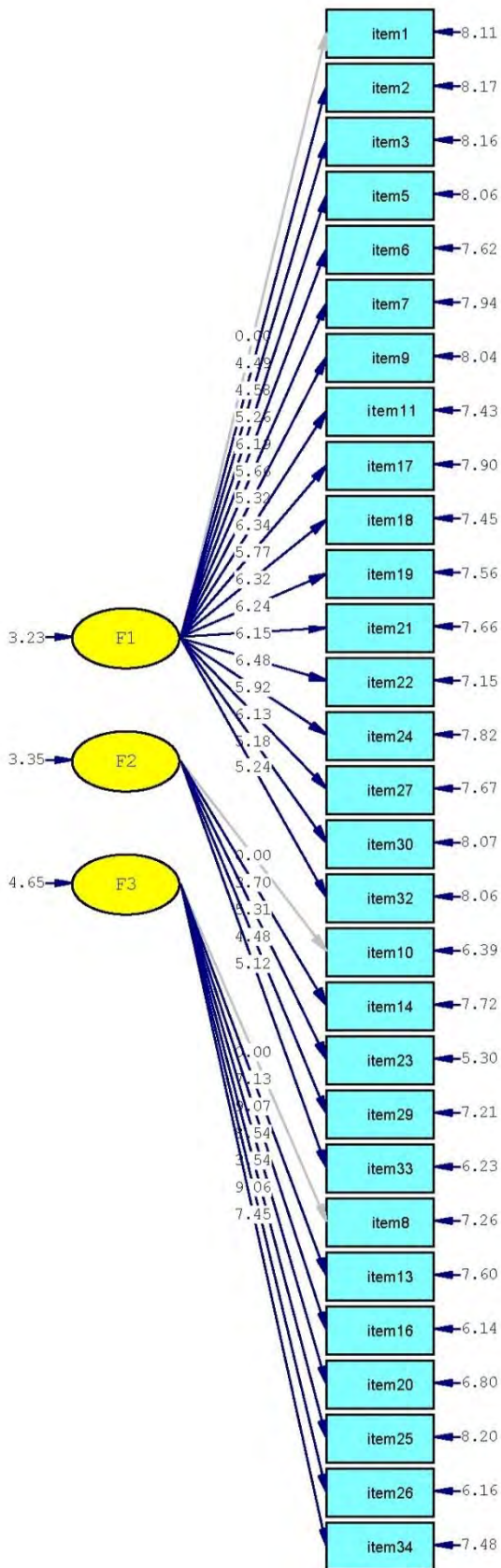


Figure 3 Path diagram containing the CFA t values for the SCES (same sample)

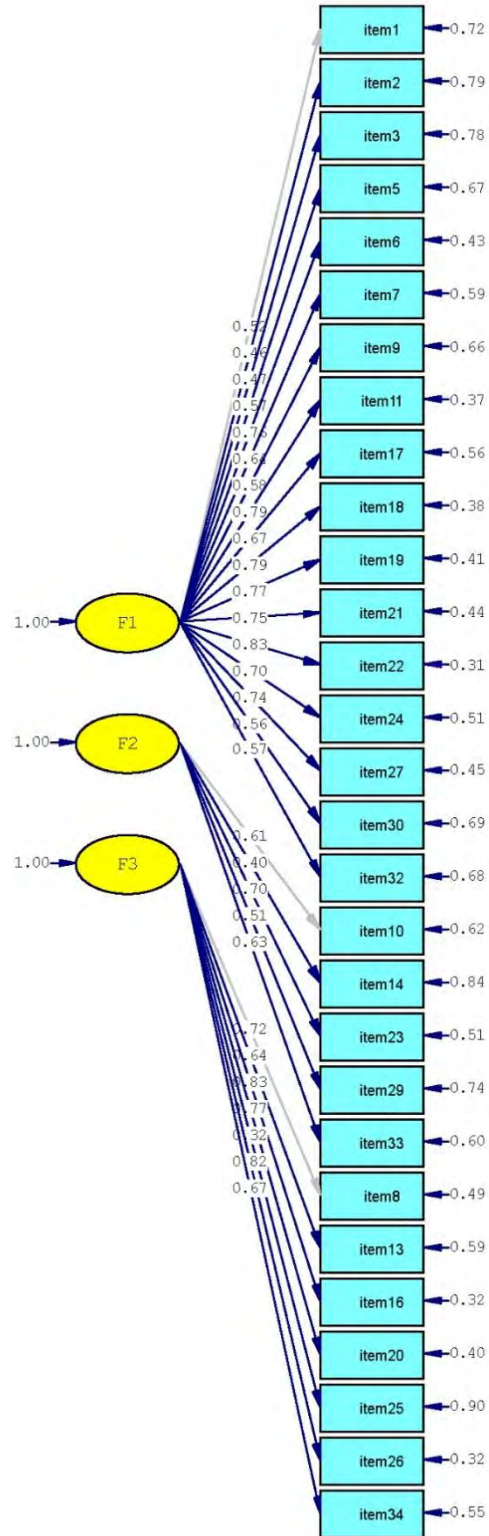


Figure 4 Path diagram containing the standardized factor loadings for the SCES (same sample)

3.3.2. Factor Analysis Findings for Different Samples

To ensure normality, the study divided the sample in two ($n = 70$ for the EFA and $n = 69$ for the CFA). The KMO and Bartlett's test results have been examined to check the suitability of the research data for factor analysis

Table 7 CFA fit indices and results for the SCES (same sample)

Fit Indexes	Acceptable limit	Perfect fit limit	Value of scale	The scale's fit decision
NFI	.90 and above	.95 and above	0.90	Acceptable
NNFI	.90 and above	.95 and above	0.93	Acceptable
IFI	.90 and above	.95 and above	0.93	Acceptable
RFI	.90 and above	.95 and above	0.89	Rejected
CFI	.95 and above	.97 and above	0.94	Rejected
GFI	.85 and above	.90 and above	0.72	Rejected
AGFI	.85 and above	.90 and above	0.67	Rejected
RMSEA	Between =.050 and =.080	Between = .000 and <.050	.080	Acceptable

Table 6 Pattern matrix values for SCES (same sample)

Item Number	Pattern Matrix Values		
	Factor 1	Factor 2	Factor 3
22	.758		
18	.749		
19	.726		
7	.725		
21	.715		
6	.708		
11	.703		
24	.703		
27	.696		
17	.685		
5	.665	.437	
9	.616		
32	.598		
30	.569		
1	.558	.429	
2	.520		
3	.512		
23		.799	
10		.703	
33		.663	
29		.510	
14		.487	
26			.730
8			.716
20			.674
13			.641
16	.327		.607
25			.510
34	.319		.508

(see Table 8). When examining Table 8, the KMO value was determined to be greater than the minimum value of .60 required for analysis; this value was statistically significant (Tabachnick & Fidell, 2007). Therefore, because the 3-factor structure was confirmed over the same sample, a structure has been obtained explaining 46.79% of the variance when performing the EFA and limiting the number of factors to three (see Table 9).

When examining Table 10, the developed scale consists of four factors and 24 items, explaining 53.48% of the

Table 8 KMO and Bartlett's test results for different samples

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.717
Bartlett's Test of Sphericity	Approx. Chi-Square	794.327
	df	276
	Sig.	.000

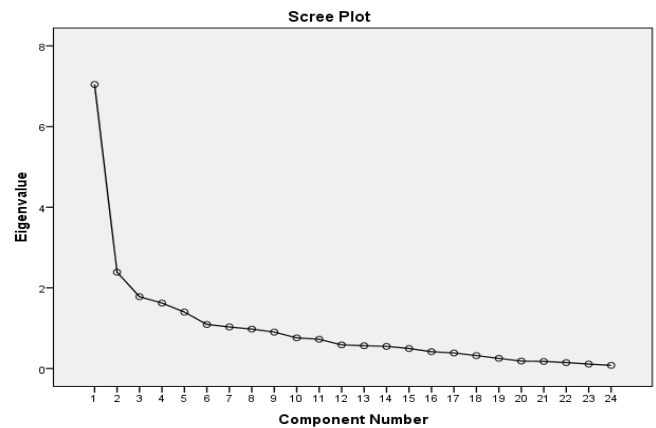


Figure 5 Scree plot for the different samples

variance. To verify the obtained structure, CFA has been performed on different samples. The number of samples required to perform CFA has been specified as 37.40, and the study performed the analysis over a sample of $n = 69$ individuals. Figure 6 provides the t values obtained as a result of the CFA. When examining the t values, Items 12, 14, 23, 29, and 33 were determined to have been indicated with red arrows, thus showing these items to be problematic. In addition, when examining the path diagram showing the standardized factor loading values, items are found with values less than .30 (see Figure 6).

The t values and standardized factor loading values obtained from the CFA analysis using different samples are not in the desired range, so the data on the other model fit index was examined (Table 11). When examining the path diagram, although the χ^2 value is understood to be 481.37 ($df = 252, p = 0.00, \chi^2 / df = 1.91$) and to have the perfect fit, the other goodness-of-fit indices do not have desired values (see Table 11). During the development phase of the SCES, the scale has been determined to consist of 24 items

Table 9 Total explained variance for the SCES (different samples)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.045	29.353	29.353	7.045	29.353	29.353
2	2.389	9.952	39.306	2.389	9.952	39.306
3	1.781	7.421	46.727	1.781	7.421	46.727
4	1.623	6.761	53.487	1.623	6.761	53.487
5	1.400	5.833	59.321			
6	1.093	4.555	63.875			
7	1.031	4.295	68.170			
8	.978	4.076	72.246			
9	.903	3.761	76.007			
10	.761	3.170	79.177			
11	.725	3.022	82.199			
12	.587	2.448	84.646			
13	.566	2.360	87.007			
14	.550	2.290	89.297			
15	.496	2.066	91.363			
16	.417	1.738	93.101			
17	.386	1.606	94.707			
18	.320	1.332	96.039			
19	.253	1.054	97.093			
20	.184	.768	97.861			
21	.176	.733	98.595			
22	.147	.613	99.207			
23	.111	.465	99.672			
24	.079	.328	100.000			

Table 10 Pattern matrix values for SCES (different samples)

Item Number	Pattern Matrix Values		
	Factor 1	Factor 2	Factor 3
11	.784		
27	.776		
18	.743		
19	.742		
24	.696		
22	.684		
16	.660		
17	.636		
8	.622		.415
21	.584		
30	.578		
3	.543		
23		.771	
4		.681	
29		.633	
33		.526	
12		.464	
25			.712
14			.631
13	.359		.603
2			.346
9			
7			
6	.402		

and four factors as a result of the EFA (Figure 5); however, this structure could not be confirmed in the CFA using different samples.

3.4. Reliability Analysis Findings

The reliability coefficients scale items have been examined for the draft SCES consisting of 34 items within the scope of reliability studies, after which the validity studies began. In addition, Cronbach's alpha of reliability has been calculated for the 29 items, and the 3-factor structure was obtained and verified as a result of the validity studies (see Table 12).

The SCES was determined to explain 52.02% of the variance, consist of 29 items and three factors, and have a reliability value of .914 due to the reliability analysis. Table 13 provides the reliability coefficients for the SCES's 29 items and their level of impact on reliability upon being removed from the scale. When examining the table, all items can be said to have acceptable values (Cronbach, 1951).

4. DISCUSSION

This study has aimed to develop a valid and reliable scale for evaluating STEM clubs. When examining the literature in terms of scale studies, scale adaptation studies are noteworthy predominant (Derin, Aydin & Kirkic, 2017; Gelen, Akcay, Tiryaki & Benek, 2019; Hacıomeroglu & Bulut, 2016) as rearranging an existing scale concerning

Table 11 CFA fit indices and results for the SCES (different samples)

Fit Indexes	Acceptable limit	Perfect fit limit	Value of scale	The scale's fit decision
NFI	.90 and above	.95 and above	.80	Rejected
NNFI	.90 and above	.95 and above	.87	Rejected
IFI	.90 and above	.95 and above	.88	Rejected
RFI	.90 and above	.95 and above	.78	Rejected
CFI	.95 and above	.97 and above	.88	Rejected
GFI	.85 and above	.90 and above	.63	Rejected
AGFI	.85 and above	.90 and above	.56	Rejected
RMSEA	Between =.050 and =.080	Between = .000 and <.050	.12	Rejected

Table 12 Reliability of the factors and SCES

Dimensions	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
SCES	.914	.928	29
Teacher	.922	-	17
Planning and Execution	.698	-	5
Student	.830	-	7

Table 13 Reliability analysis results of SCES items

Item Number	Corrected Item Total Correlation	Cronbach's Alpha if Item Deleted
1	.569	.910
2	.452	.912
3	.436	.912
5	.578	.910
6	.709	.908
7	.551	.911
8	.505	.911
9	.559	.911
10	.360	.914
11	.719	.908
13	.503	.911
14	.290	.917
16	.683	.910
17	.542	.911
18	.690	.909
19	.652	.910
20	.596	.910
21	.650	.909
22	.731	.908
23	.247	.916
24	.561	.911
25	.217	.918
26	.640	.910
27	.616	.910
29	.260	.917
30	.543	.911
32	.496	.912
33	.369	.914
34	.612	.910

a different culture and language rather than developing a new scale is thought to save both time and money (Oner,

2008). This situation leads to the absence of scales for specific areas (Acar-Güvendir & Özer-Özkan, 2015). The current study has developed a scale to evaluate the work of teachers who carry out STEM club activities by choosing a subject area that has not been previously studied. In addition, the increase in scale development and adaptation studies in the literature related to STEM and science teachers draws attention (Corlu et al., 2014; Derin, Aydin & Kirkic, 2017; Hacıomeroglu & Bulut, 2016; Unlu, Dokme, & Veli, 2016). When examining these studies, they are seen to mostly focus on affective factors such as attitudes (Derin, Aydin & Kirkic, 2017), motivation (Donmez, 2020), and awareness (Buyruk & Korkmaz, 2016; Cevik, 2017). No scale study was found regarding obtaining teachers' opinions toward STEM clubs or the work carried out by these clubs. The scale developed in this respect is thought to fill a gap in the field and help organize STEM club activities.

The current research has developed the STEM Club Evaluation Scale for teachers who conduct STEM club activities. This study followed DeVellis' (2014) 8-step scale development method. In addition, the scale items were created based on the constructivist theory, the philosophical approach on which STEM education is based, and upon which the scale focuses. Although scales having a theoretical basis is important, a limited number of studies are seen to have been shaped within the framework of a theoretical structure in the literature (Kizilay, Yamak & Kavak, 2019). Similarly, Kizilay, Yamak & Kavak (2019) study carried out the scale development process according to DeVellis' (2014) scale development steps and took into account the motivation-based ARCS model.

A theoretical framework was formed to ensure the scale's content validity, a literature review was conducted, expert opinions were taken, and item difficulty and

discrimination indexes were calculated. As a result of these calculations, items were seen to have low item discrimination indexes. However, these items were not immediately removed from the scale until the other analysis results were examined.

After enabling the scale's content validity, its construct validity was checked by performing EFA and CFA. When examining scale development studies in the field of science education, studies are found where only EFAs have been performed and no CFAs (Cermik & Kara, 2020; Firdaus, Subchan & Narulita, 2020), where EFA and CFA were conducted over the same sample (Kizilay, Yamak & Kavak, 2019; Pedaste, Baucal & Reisenbuk, 2021), and where EFA and CFA were performed over different samples (Akkus, 2019; Burak & Gultekin, 2021; Fidan & Tuncel, 2021; Yildirim & Sahin-Topalcengiz, 2018). When looking at scale development studies in this context, the structure EFA reveals is seen should be verified using CFA. On the other hand, a disagreement exists in the literature regarding whether CFA should be performed over the same sample as in the EFA or over a different one. Therefore, the current study has conducted its EFA and CFA using both the same sample and different samples to clarify the confusion in the literature. As a result of the EFA performed over the same sample, a structure was obtained consisting of three factors and 29 items explaining 52.02% of the variance, and the CFA confirmed this structure. An analysis that explains 50-75% of the total variance is considered a valid analysis (Beavers et al., 2013). The developed scale can be said to be valid in this respect. As a result of the EFA performed on different samples, a structure was obtained consisting of four factors and 24 items explaining 53.48% of the variance; however, CFA did not confirm this structure. Although the literature argues that a different sample should be used for CFA and EFA, in light of the results the current study obtained, the same sample should at least be split in half and analyses made over the two groups. Analyzing the same data set with different software has also been said to be sufficient for confirming the factor structure (Yaslioglu, 2017). The CFA did not confirm the structure obtained in the current study as a result of the EFA made by creating two groups and providing normality for both groups as suggested by the literature. This result can be considered a reference for future scale development studies on the point of factor analysis validation.

The factor loads for the items from the three factors have values between .487 and .799. For an item to be included under any factor, it must have a value of at least .30. We calculated item discrimination and difficulty indices before factor analysis. The discrimination index of item 15 was low and negative. As a result of item reliability analysis, we found the reliability of items 15, 28, and 31 negative. Karakaplan & Yildiz (2010) also removed four items that harmed reliability while calculating the Alpha value in their

scale development studies. Therefore, we did not include these three items in the factor analysis. Since the reliability of item 4 was below .20, we did not include it in the factor analysis. The item's correlation coefficient does not affect the item reliability coefficient (Cronbach's alpha) value, so the item-total score correlation value should be at least 0.20 (Tavsancil, 2002). In the exploratory factor analysis, item 12 was excluded from the scale because it did not fit under any factor. Similarly, Yolagiden & Bektas (2021), in their study of "Developing an Entrepreneurship Scale for Science Course", removed item 21 from the scale because it did not fall under any factor.

The factor loading values for the items on the SCES have been considered good in this respect. The factors that were identified by receiving expert opinions are the teacher (Factor 1), planning and implementation (Factor 2), and student (Factor 3). Noteworthy, a similar scale development study in the literature found the following factor names: STEM's effect on the student, STEM's effect on the lesson, and STEM's effect on the teacher (Cevik, 2017). For example, the "I have sufficient knowledge about STEM education." item was asked to determine teacher proficiency regarding STEM club practices. For this reason, it is foreseen that this item will be included in the teacher dimension. Although the item "The STEM club plans I use are not suitable for the student level." mentions STEM club plans, it has been evaluated in the student dimension since it emphasizes the suitability of the plans for the student level. Although the item "The time allocated for STEM club activities carried out in my school is not enough." It seems to be for the teacher and has been included in the planning and implementation dimension because it stems from the structure of the social club rules.

Within the scope of reliability studies, the reliability coefficients from the draft scale first consisted of 34 items examined each item individually. The validity studies began after examining the reliability coefficients. Cronbach's alpha of reliability was calculated for the 29-item 3-factor scale and was determined and verified as a result of the validity studies. As a result of the reliability analysis, the SPSS was concluded to explain 52.02% of the variance, consisted of 29 items and three factors, and have a reliability value of .914. Yılmaz & Cavas (2007) study calculated the reliability of each item by subjecting them to separate reliability analyses in the program SPSS; the reliabilities calculated for each factor of the scale were seen to range from .54 to .85. When examining the reliabilities for each item in this study, the values were determined to vary between .217 and .719. In addition, the reliability was calculated for each factor. For example, the reliability for the factor of the teacher is .922. The reliability factor for planning and implementation was calculated as .698, and student factor was calculated as .830. Thus, the SCES is concluded to have high overall reliability for each factor and item (Cronbach, 1951). In this context, knowing

whether the obtained results are reliable using the whole scale and for each factor was desired.

5. CONCLUSION

As a result of the validity checks, we determined that 29 items and a three-factor structure explained 52% of the variance. As a result of the reliability checks, we calculated the Cronbach's alpha reliability as .92. As a result, a valid and reliable measurement tool has been developed by which program practitioners and researchers can measure the effectiveness and efficiency of STEM clubs.

SUGGESTIONS

- The SCES can determine the effectiveness and efficiency of STEM clubs at the provincial, district, and school levels.
- The developed scale can be used as a data collection tool to create framework plans for STEM club studies.
- The SCES can be applied to different levels of education by making it suitable in terms of language and intelligibility.
- The SCES can be used as a data collection tool in studies conducted in different socioeconomic and geographical regions regarding academic achievement, gender, and family.

REFERENCES

- Acar-Güvendir, M., & Özer-Özkan, Y. (2015). The examination of scale development and scale adaptation articles published in Turkish academic journals on education. *Electronic Journal of Social Science*, 14(52), 23-33. <https://doi.org/10.17755/esosder.54872>
- Afterschool Alliance. (2015). Full STEM Ahead: Afterschool Programs Step Up as Key Partners in STEM Education. Washington, D.C. Retrieved from <http://www.afterschoolalliance.org/AA3PM/>.
- Akgunduz, D., Aydeniz, M., Cakmakci, G., Cavas, B., Corlu, M. S., Oner, T., & Ozdemir, S. (2015). *STEM education Turkey report*. Istanbul: Scala Press.
- Akkus, A. (2019). Developing a Scale to Measure Students' Attitudes toward Science. *International Journal of Assessment Tools in Education*, 6(4), 706-720. <https://doi.org/10.21449/ijate.548516>
- Altunel, M. (2018). STEM education and Turkey: opportunities and risks. *Seta Perspective*, 207, 1-7.
- Ayers, K. A., Wade-Jaimes, K., Wang, L., Pennella, R. A., & Pounds, S. B. (2020). The St. Jude STEM Clubs: An After-school STEM Club for Upper Elementary School Students in Memphis, TN. *Journal of STEM Outreach*, 3(1), 1-26. <https://doi.org/10.15695/jstem/v3i1.13>
- Baldrige, A., Nutt, A., Vaughn, M., Hartley-Lewis, C., & Amos, A. (2009). *The STEM Club at Marietta High School*. ASEE Southeast Section Conference.
- Baran, E., Bilici, S. C., 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. <http://dx.doi.org/10.18404/ijemst.71338>
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, 18(1), 6. <https://doi.org/10.7275/qv2q-rk76>
- Bell, P., Lewenstein, B., Shouse, A.W., & Feder, M.A. (2009). *Learning science in informal environments: People, places, and pursuits*. Washington, DC: The National Academies Press.
- Birturk, H. (2015). *The functionality of social clubs in schools*. Unpublished master's thesis. Yeditepe University Institute of Educational Sciences, Istanbul.
- Blanchard, M. R., Hoyle, K. S., & Gutierrez, K. S. (2017). How to start a STEM club. *Science Scope*, 41(3), 88-94.
- Burak, D., & Gultekin, M. (2021). Verbal-Visual Learning Styles Scale: Developing a Scale for Primary School Students. *International Journal on Social and Education Sciences*, 3(2), 287-303. <https://doi.org/10.46328/ijones.171>
- Buyruk, B., & Korkmaz, Ö. (2016). STEM Awareness Scale (SAS): Validity and Reliability Study. *Journal of Turkish Science Education*, 11(1), 3-23. <https://doi.org/10.12973/tused.10179a>
- Buyukuzturk, S., Kilic-Cakmak, E., Akgun, O. E., Karadeniz, S. & Demirel, F. (2016). *Scientific research methods (22nd Edition)*. Ankara: Pegem Academy.
- Bybee, R.W. (2001). Achieving scientific literacy: Strategies for ensuring that free choice science education complements national formal science education efforts. In J.H. Falk (Ed.), *Free choice education: How we learn science outside of school* (pp. 44–63).
- Cermik, H., & Kara, I. (2020). Physics course attitudes scale for high school students: a validity and reliability study. *International Journal of Assessment Tools in Education*, 7(1), 62-72. <https://doi.org/10.21449/ijate.693211>
- Cevik, M. (2017). A study of STEM Awareness Scale development for high school teachers. *Journal of Human Sciences*, 14(3), 2436-2452. DOI:10.14687/jhs.v14i3.4673
- Corlu, M. S., Capraro, R. M., & Capraro, M. M. (2014). Introducing STEM education: Implications for educating our teachers in the age of innovation. *Education and Science*, 39(171), 74-85. <http://hdl.handle.net/11693/13203>
- Coskun, T. K., Alakurt, T., & Yilmaz, B. (2020). STEM Education from the Perspective of Information Technologies Teachers. *Abant İzzet Baysal University Journal of the Faculty of Education*, 20(2), 820-836. <https://doi.org/10.17240/aibuefd.2020.-536856>
- Creswell, J. W. (2017). *Research design qualitative, quantitative, and mixed-method studies* (3rd edition) (S. B. Demir, Trans. Ed.). Ankara: Educating Book.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Dabney, K. P., Tai, R. H., Almarode, J. T., Miller-Friedmann, J. L., Sonnett, G., Sadler, P. M., & Hazari, Z. (2012). Out-of-school time science activities and their association with a career interest in STEM. *International Journal of Science Education, Part B*, 2(1), 63-79. <https://doi.org/10.1080/21548455.2011.629455>
- Derin, G., Aydin, E., & Kirkic, K. A. (2017). A Scale on the attitudes towards STEM education. *El-Cezeri Journal of Science and Engineering*, 4(3), 547-559. <https://doi.org/10.31202/ecjse.336550>
- DeVellis, R. F. (2014). Scale development: Theory and applications. <https://doi.org/10.14689/ejer.2016.63.2>
- Donmez, I. (2020). Adaptation of STEM Motivation Scale into Turkish: Validity and Reliability Study. *YYU Journal of Education Faculty*, 17(1), 486-510. <https://doi.org/10.33711/yyuefd.693825>
- Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Eroglu, S., & Bektas, O. (2016). Ideas of Science Teachers took STEM Education about STEM-based Activities. *Journal of Qualitative Research in Education*, 4(3), 43-67. <https://doi.org/10.14689/issn.2148-2624.1.4c3s3m>
- Faber, M., Unfried, A., Wiebe, E. N., Corn, J., Townsend, L. W., & Collins, T. L. (2013, June). Student attitudes toward STEM: The development of the upper elementary school and middle/high school student surveys. In *2013 ASEE Annual Conference & Exposition* (pp. 23-1094).

- Ferrara, M., Mason, H., Wee, B., Rorrer, R., Jacobson, M., & Gallagher, D. (2017). Enriching undergraduate experiences with outreach in school STEM clubs.
- Fidan, M., & Tuncel, M. (2021). Developing A Self-Efficacy Scale Toward Physics Subjects For Lower-Secondary School Students. *Journal of Baltic Science Education*, 20(1), 38. <https://doi.org/10.33225/jbse/21.20.38>
- Firdaus, F., Subchan, W., & Narulita, E. (2020). Developing STEM-based TGT learning model to improve students' process skills. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 6(3), 413-422. <https://doi.org/10.22219/jpbi.v6i3.12249>
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). How to design and evaluate research in education.
- Fraenkel, J., & Wallen, N. (1996). *Validity and reliability. How to design and research in education*. New York: McGraw-Hill, INC, 3, 153-171.
- Gabrielson, E. A., Strachan, J. L., Warner, M. T., & LaFleche, S. H. (2009). Evaluating the London Science Museum's Activity Boxes at UK STEM Clubs.
- Gelen, B., Akcay, B., Tiryaki, A., & Benek, I. (2019). Pre-Service Science Teachers' Self-Efficacy toward Science, Technology, Engineering, Mathematics (STEM) Survey: An Adaptation to Turkish, Validity and Reliability Study. *Journal of Theory and Practice in Education*, 15(1), 88-107. <https://doi.org/10.17244/eku.395204>
- George, D., & Mallery, P. (2016). *IBM SPSS statistics 26 step by step: A simple guide and reference*. (14th ed.). Routledge.
- Gonsalves, A., Rahm, J., & Carvalho, A. (2013). "We could think of things that could be science": Girls' re-figuring of science in an out-of-school-time club. *Journal of Research in Science Teaching*, 50(9), 1068-1097. <https://doi.org/10.1002/tea.21105>
- Gottfried, M. A., & Williams, D. (2013). STEM club participation and STEM schooling outcomes. *Education Policy Analysis Archives*, 21, 79. DOI:10.14507/epaa.v21n79.2013
- Guzey, S. S., Harwell, M., & Moore, T. (2014). Development of an instrument to assess attitudes toward science, technology, engineering, and mathematics (STEM). *School Science and Mathematics*, 114(6), 271-279. <https://doi.org/10.1111/ssm.12077>
- Hacıomeroglu, G., & Bulut, A. S. (2016). Integrative STEM teaching intention questionnaire: a validity and reliability study of the Turkish form. *Journal of Theory and Practice in Education*, 12(3), 654-669. <http://eku.comu.edu.tr/article/view/5000176286/5000164803>
- Hair, J. F., Black, W. C., Tatham, R. L. & Anderson, R. E. (2019). *Multivariate data analysis* (Eighth Edition). Cengage Learning EMEA.
- Johnson, B., & Christensen, L. (2014). *Educational research: Quantitative, qualitative and mixed approaches* (Trans. Ed. Demir, S. B.). Ankara: Egiten Book.
- Jöreskog, K., & Sörbom, D. (1993). *Lisrel 8: structural equation modeling with the simplis command language*. Lincolnwood: Scientific Software International, Inc.
- Kalkan, C., & Eroglu, S. (2017). Designing sample activities based on STEM materials for gifted/talented students in support education rooms. *Journal of Gifted Education and Creativity*, 4(2), 36-46. <https://dergipark.org.tr/tr/pub/jgedc/issue/38702/449432>
- Karadogan, S. (2016). In education-school learning practices and daily classroom problems. In Academic Evaluations and Solution Suggestions for Educational Problems in Turkey-1, (Ed. R Aksu), *Publisher Maya Academy*, 47-84.
- Karakaplan, S., & Yildiz, H. (2010). A Study On Developing A Postpartum Comfort Questionnaire. *Maltepe University Journal of Nursing Science and Art*, 3(1), 55-65.
- Kizilay, E., Yamak, H., & Kavak, N. (2019). Motivation Scale for STEM Fields. *Journal of Computer and Education Research*, 7(14), 540-557. <https://doi.org/10.18009/jcer.617514>
- Lipuma, J., Bukiet, B. G., & Leon, C. (2021). *Hands-on Developmental Playbook for STEM Clubs in Elementary Schools*. STEM for Success Resources. 3. <https://digitalcommons.njit.edu/stemresources/3>
- Mahoney, J. L., Parente, M. E., & Lord, H. (2007). After-school program engagement: Links to child competence and program quality and content. *The Elementary School Journal*, 107(4), 385-404.
- 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. <https://doi.org/10.1177%2F1069072713515427>
- Ministry of National Education. (2016). STEM Education Report. http://yegitek.meb.gov.tr/STEM_Egitimi_Raporu.pdf, (Access date: 10 September 2021).
- Ministry of National Education. (2018). Curriculum Monitoring and Evaluation System-Curriculums. <http://mufredat.meb.gov.tr/Programlar.aspx>, (Access date: 10 September 2021).
- MoNE, (2017). MoNE Educational Institutions Social Activities Regulation.
- National Research Council (2015). *Identifying and supporting productive STEM programs in out-of-school settings*. National Academies Press.
- Oner, N. (2008). Examples of psychological tests used in Turkey: A reference (extended 2nd edition). Istanbul: Bogazici University Publishing.
- Pallant, J. (2016). *SPSS user guide Step-by-step data analysis with SPSS*. (S. Balci & B. Ahi, Trans.). Ankara: Ani Publishing.
- Pedaste, M., Baucal, A., & Reisenbuk, E. (2021). Towards a science inquiry test in primary education: development of items and scales. *International Journal of STEM Education*, 8(1), 1-19.
- Polat, B. S. (2017). *Investigation of Opinions of Administrators, Teachers and Students About the Effectiveness of Social Club Activities in Secondary Schools*. Unpublished master's thesis. Ataturk University Institute of Education Sciences, Erzurum.
- Qasem, M. A. N., & Gul, S. B. A. (2014). *Effect of items direction (positive or negative) on the factorial construction and criterion-related validity in Likert scale*. *Khazar Journal of Humanities and Social Sciences*, 17(3), 77-84. <http://hdl.handle.net/20.500.12323/3240>
- Robelen, E. (2011). New STEM schools target underrepresented groups. *Education Week*, 31(1), 18-19.
- Sahin, A. (2013). STEM clubs and science fair competitions: Effects on post-secondary matriculation. *Journal of STEM Education: Innovations and Research*, 14(1), 5-11.
- Sahin, A., Ayar, M. C. & Adigüzel, T. (2014). STEM Related After-School Program Activities and Associated Outcomes on Student Learning. *Educational Sciences: Theory and Practice*, 14(1), 309-322. <http://dx.doi.org/10.12738/estp.2014.1.1876>
- Secer, İ. (2017). *Practical Data Analysis with Spss and Lisrel: Analysis and Reporting* (3rd ed). Ankara: Ani Publishing.
- Siew, N. M., Amir, N., & Chong, C. L. (2015). The perceptions of pre-service and in-service teachers regarding a project-based STEM approach to teaching science. *SpringerPlus*, 4(1), 8. <http://dx.doi.org/10.1186/2193-1801-4-8>
- Straw, A. D., Branson, K., Neumann, T. R., & Dickinson, M. H. (2011). Multi-camera real-time three-dimensional tracking of multiple flying animals. *Journal of The Royal Society Interface*, 8(56), 395-409.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.
- Tavsancil, E. (2002). *Measuring attitudes and data analysis with SPSS*. Ankara: Nobel Publishing. Turgut, M. F., & Baykul, Y. (2015). *Measurement and evaluation in education* (7th Edition). Ankara: PegemA Publishing.
- TUSIAD (2014). A research on the demands and expectations of the workforce trained in STEM (Science, Technology, Engineering and Mathematics, Science, Technology, Engineering, Mathematics). TUSIAD.
- Unlu, Z. K., Dokme, I., & Veli, U. N. L. U. (2016). Adaptation of the science, technology, engineering, and mathematics career interest survey (STEM-CIS) into Turkish. *Eurasian Journal of Educational Research*, 16(63), 21-36 <http://dx.doi.org/10.14689/ejer.2016.63.2>
- Vural, C. (2018). *The Determination of School Principals' and Teachers' Attitude Towards Educational School Clubs*. Unpublished Master's

- thesis. İstanbul Sabahattin Zaim University, Social Sciences Institute, İstanbul.
- Yaslioglu, M. M. (2017). Factor analysis and validity in social sciences: application of exploratory and confirmatory factor analyses. *Istanbul University Journal of the School of Business*, 46, 74-85.
- Yildirim, B., & Sahin-Topalcengiz, E. (2018). STEM Pedagogical content knowledge scale (STEMPCK): A validity and reliability study. *Journal of Baltic Science Education*, 20(1), 38-49.
- Yılmaz, H., & Çavaş, P. H. (2007). Reliability and validity study of the Students' Motivation toward Science Learning (SMTSL) Questionnaire. *Elementary education online*, 6(3).
- Yolagiden, C., & Bektas, O. (2021). Development of entrepreneurship scale for science lesson: a validity and reliability study. *Afyon Kocatepe University Journal of Social Sciences*, 23(4), 1349-1365. <https://doi.org/10.32709/akusosbil.903893>
- Zengin, N., Kaya, G., & Pektaş, M. (2020). STEM temelli arařtırmalarda kullanılan ölçme ve deęerlendirme yöntemlerinin incelenmesi. *Gazi University Journal of Gazi Educational Faculty (GUJGEF)*, 40(2), 329-355.