Astronomy Teaching Self-Efficacy Belief Scale: The Validity and Reliability Study

Filiz Demirci[1] & Cengiz Ozyurek[1]

[1] Department of Science Education, Ordu University, Ordu, Turkey

Correspondence: Filiz Demirci, Department of Science Education, Ordu University, Ordu, Turkey. E-mail: filizdemirci@odu.edu.tr

Abstract

The purpose of this study is to develop a reliable and safe scale for determining the self-efficacy levels of science teachers in the teaching of astronomy subjects. The study used a survey approach, which is a qualitative research method. The study was conducted with a total of 106 science teachers working in the secondary schools of Ordu city centre and the surrounding towns during the academic year 2016-2017. While forming the item pool of the scale, scale development studies within the context of teacher self-efficacy and the special field competencies of science and technology teachers determined by MOE (2008) was used. In addition, the compositions written by eight science teachers outside the study group about the teaching of astronomy were also used for item pool. For the content validity of the scale, an expert opinion form was prepared to assess the content validity rate and kappa coefficient of agreement, and this was presented to six faculty members in the science teaching department. The construct validity of the scale was investigated via exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The results of EFA showed that the scale construct included a total of three factors and 13 questions, and explained 70.60% of the total variance. CFA results showed that the chi-squared value and the degrees-of-freedom rates (χ^2/sd = 1.67) were perfect, and the other fit indices showed a good fit (GFI = 0.86, CFI = 0.94, NNFI = 0.92, IFI = 0.94, SRMR = 0.08 and RMSEA = 0.06). The results of the reliability analysis showed that the Cronbach's alpha reliability coefficient was 0.84 for the whole scale, 0.90 for "student outcomes through astronomy teaching" factor, and 0.83 for both "astronomy teaching strategies" factor and "difficulty in astronomy teaching" factor. In conclusion, the results obtained showed that "Astronomy Teaching Self-Efficacy Belief Scale" can be used as a valid and reliable assessment instrument.

Keywords: astronomy teaching, scale development, self-efficacy belief, teacher self-efficacy

1. Introduction

In terms of pupils' learning and development, personal beliefs about their teaching behavior as well as professional knowledge and skills of teachers are important. Teachers' beliefs about teaching competences are closely related to performances in the teaching process. Self-efficacy belief is one of the personal beliefs that directly influence the performance that teachers will exhibit to attain certain goals.

Self-efficacy is defined as "one's beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). Bandura (1977) mentioned the self-efficacy concept for the first time in his social cognitive learning theory based on a behaviour change mechanism. On the basis of social-cognitive learning theory, teacher self-efficacy beliefs are consistently associated with student outcomes and effective teaching behaviors (Henson, 2001).

Bandura bases his theoretical self-efficacy model on two basic components: "outcome expectancy (is defined as a person's estimate that a given behaviour will lead to certain outcomes) and efficacy expectation (is the conviction that one can successfully execute the behaviour required to produce the outcomes)" (Bandura, 1977, p. 193). These constructs proposed by Bandura on self-efficacy (efficacy expectation and outcome expectation) form a theoretical framework in studies about teacher self-efficacy.

According to Tschannen-Moran, Woolfolk-Hoy, & Hoy (1998), teacher self-efficacy is the teacher's belief about his or her ability to organize and carry out the acts required to successfully perform a specific teaching task in a particular field. According to Ross & Bruce (2007), teacher self-efficacy is a teacher's expectancy that he or she

can make the students learn. According to this definition, it can be said that teachers' self-efficacy beliefs about a specific subject will be reflected in their classroom behaviours. There are many studies in the literature which show that there is a direct association between teachers' self-efficacy beliefs and student outcomes such as achievement, attitude, motivation, self-efficacy, etc. There are many studies showing a strong association between teachers' self-efficacy beliefs and student outcomes (achievement: Ross & Bruce, 2007, motivation: Midgley et al., 1989; Ford, 2012 and attitude: Show-Alter, 2005). Friedman & Kass (2002) stated that teachers' self-efficacy beliefs are fundamentally associated with teaching performance. In addition, according to the results of class observations conducted by Gibson & Dembo (1984), there are differences between the classroom behaviours (the amount of time allocated to academically oriented and whole-class instruction, teachers' lack of persistence in cases of failure, etc.) of teachers with low and high self-efficacy beliefs.

There is a close association between teachers' self-efficacy beliefs and the strategies, methods and classroom practices they use during educational activities (Azar, 2010). In addition, it has been stated that teacher self-efficacy belief is an important construct which can influence teachers' professional development efficiency and can also influence how a new curriculum will be applied (Blonder et al., 2014)

Individuals' self-efficacy beliefs about a subject influence how they think, how they feel, how they act and how they motivate themselves in every field of their lives (Bandura, 1997). Studies show that teachers with high self-efficacy demonstrate more effective teaching behaviours, are more successful in teaching (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998), allocate more time to academically oriented activities in their classroom, help students who have difficulties in learning to succeed and motivate such students (Gibson & Dembo, 1994), show more consistent behaviours in cases of failure (Protheroe, 2008) and are more prone to use questioning-based teaching methods and cooperative learning activities (Gavora, 2010). On the other hand, it has been observed that teachers with low self-efficacy beliefs spend more time on non-academic pursuits, criticize their students more negatively when they fail and when they ask a question directed to a particular student, tend to ask another student without waiting for the first student to find the right answer, or else they just ask another question (lack of persistence) (Gibson & Dembo, 1984).

A large number of researchers from the end of the 1970s to the present day have used Bandura's (1977) social cognitive theory to explain differences in teachers' practices and students' success (Roberts & Henson, 2000). For this reason, many researchers have measured teachers' self-efficacy based on the two-factor theoretical model of self-efficacy (self-efficacy and outcome expectancy) put forward by Bandura. One of the first measurement instruments to measure the self-efficacy beliefs of teachers, the "Teacher Efficacy Scale" (TES) was developed by Gibson & Dembo (1984). This scale, which was developed to measure the teacher self-efficacy beliefs of primary school teachers, consists of a total of 16 items and two sub-factors (factor 1 is personal teaching efficacy and factor 2 is general teaching efficacy or outcome expectancy). The Cronbach's alpha reliability coefficient is 0.79 for the whole scale, 0.78 for the personal teaching efficacy factor and 0.75 for the teaching efficacy factor.

Riggs & Enochs (1990) tried to show that teacher self-efficacy is a special field of study and a special situation. Within this context, they discussed the two dimensions of teacher self-efficacy in the literature (teaching efficacy or outcome expectancy and personal teaching efficacy or self-efficacy). They developed the "Science Teaching Efficacy Beliefs Instrument" (STEBI—Form A) based on Gibson & Dembo's (1984) two-factor TES. This instrument, which was developed to measure the self-efficacy beliefs of primary school teachers towards science education, included a total of 25 items and two factors. The Cronbach's alpha reliability coefficients were 0.92 for personal science teaching efficacy belief and 0.77 for science teaching outcome expectancy.

Bandura (1981) stated that the self-efficacy belief is a special situation rather than a general construct (cited from Riggs & Enochs, 1990). Pajares (1996) stated that self-efficacy scales which cover a general subject have low predictability for behaviours which belong to a specific field. Especially in the case of primary school teachers, a measurement instrument which addresses a specific subject can be more informative about the teacher's self-efficacy belief (Riggs & Enochs, 1990). Within this context, scales have also been developed or adapted to determine the self-efficacy of teachers regarding the more specific field of science teaching (physics teaching: Barros et al., 2010, chemistry teaching: Morgil et al., 2004, biology teaching: Kiremit, 2006, environmental education: Sia, 1992; Özdemir et al., 2009 and astronomy teaching: Güneş, 2010). For instance, in their study, Barros et al. (2010) developed a scale which included a total of 18 items and two sub-factors to measure the physics instruction self-efficacy beliefs of secondary education physics teachers in Brazil. In their scale development study, Barros et al. (2010) adapted the items of Woolfolk & Hoy (1990) and Riggs & Enochs (1990). The personal efficacy belief of physics teachers, which is one of the sub-factors of the developed scale, includes special items about science instruction and other subjects relating to the teaching of physics (for

example, experimentation, conceptual structure and mathematical formalism). The Cronbach's alpha reliability coefficient of this factor, which included a total of nine items, was 0.78. General efficacy belief in physics teaching, which is the other sub-factor of the scale, includes general items which aim to measure the self-efficacy beliefs of teachers with respect to physics instruction. The Cronbach's alpha reliability coefficient of this factor, which included a total of nine items, was 0.61.

Tschannen-Moran & Woolfolk-Hoy (2001) developed the "Teachers' Sense of Teacher Efficacy Scale" TSTES (also called the Ohio State Teacher Efficacy Scale), which has two forms: a long and a short form. Both scale forms consist of three factors: instructional strategies, classroom management and student engagement. The TSTES short form consists of 12 questions, with a Cronbach's alpha reliability coefficient of 0.90, while the long form consists of 24 questions, with a Cronbach's alpha reliability coefficient of 0.94.

Although there are many scale development studies in the literature about the science teaching self-efficacy beliefs of teachers or prospective teachers (Riggs & Enochs, 1990; Roberts & Henson, 2000, Ritter et al., 2001; Kaya et al., 2014), no measurement instrument was found which aimed to determine the self-efficacy beliefs of in-service science teachers about the teaching of astronomy subjects. However, Güneş (2010) developed the "Astronomy Instruction Self-Efficacy Belief Scale" to determine the self-efficacy beliefs of pre-service teachers about astronomy instruction. In his study, Güneş (2010) changed the word "science" to the word "astronomy" in the items of the "Science Teaching Self-Efficacy Belief Scale" developed by Riggs & Enochs (1990) and adapted into Turkish by Özkan et al. (2002). The validity and reliability studies on the scale (which consisted of a total of 23 items and two factors) were conducted on prospective teachers of science and social studies. The Cronbach's alpha reliability of the scale and its sub-factors were 0.80 for the whole scale, 0.87 for astronomy instruction personal self-efficacy and 0.78 for astronomy instruction expectations.

In conclusion, it can be seen from the scale development studies in the literature (Gibson, 1983) that teacher self-efficacy belief has a multidimensional structure consisting of at least the two sub-dimensions of personal teaching efficacy and general teaching efficacy.

1.1 The Rationale and the Purpose of the Study

It has been found that in subjects related to astronomy, which is a specific field in science lessons, students have too many basic misconceptions or they have low conceptual comprehension (Baxter, 1989; Zeilik et al., 1997; Diakidoy & Kendeou, 2001; Agan, 2004; Peña & Quílez, 2001; Plummer & Zahm, 2010). When it is considered that the most important factor influencing student outcomes is the effectiveness of the teacher's classroom behaviours (Rodger et al., 2007; Ali-Shah, 2009) and that self-efficacy belief is a construct which directly influences teachers' instructional behaviours (Bandura, 1977). It can be concluded that the main factor underlying this situation may be the low self-efficacy beliefs of teachers regarding astronomy subjects. Thus, in order to improve science teachers' self-efficacy beliefs about the teaching of astronomy subjects, their existing self-efficacy belief levels should first be determined.

When the literature is reviewed, it is found that the astronomy instruction self-efficacy belief scale (Güneş, 2010) is adapted from a widely used scale which was developed to discover the self-efficacy beliefs of prospective teachers. However, the fact that this scale has cultural differences means that there is a need for a specific assessment instrument to discover the self-efficacy beliefs of science teachers about astronomy teaching. For these reasons, this study aims to develop a valid and reliable assessment instrument which measures the self-efficacy beliefs of science teachers in the teaching of astronomy subjects. In addition, it was also taking into consideration the special field competencies of science and technology teachers determined by MOE (2008). Thus, it is hoped that this scale will be used as an assessment instrument for discovering the self-efficacy belief levels of teachers.

2. Method

In this study, survey research, one of the qualitative research methods, was used (Fraenkel, Wallen & Hyun, 2012). The survey method was used since the research aims to define the self-efficacy beliefs of science teachers via as many individuals as possible.

2.1 Research Group

A total of 106 science teachers (57 women and 49 men) working in the province and towns of Ordu during the 2016-2017 academic year participated in the study voluntarily. The sample for the research was chosen using the convenience sampling method, which is easy to access and practical (Fraenkel et al., 2012).

2.2 Development Stages of Astronomy Teaching Self-Efficacy Belief Scale (ATSBS)

The development process of the scale was based on the scale development steps of DeVellis (2014) and Seçer (2015).

2.2.1 Specification of the Construct to Be Measured

When the studies in the literature were examined, the existing shortcomings were found, and the present study aimed to develop a more reliable and valid self-efficacy belief scale regarding the teaching of basic astronomy subjects.

2.2.2 Forming the Item Pool

For the ATSBS to be developed, studies on self-efficacy were first researched. Scale development studies in the literature on self-efficacy beliefs in science instruction (Riggs & Enochs, 1990; Ritter et al., 2001; Friedman & Kass, 2002) were reviewed. In addition, items which specified the self-efficacy beliefs of teachers were noted by making use of the special field competencies of science and technology teachers determined by MOE (2008). Eight science teachers outside the research group were asked to write a composition on their opinions, behaviours and beliefs about the teaching of astronomy subjects. Content analysis was conducted on the teachers' expressions in their texts about self-efficacy beliefs, and added to the item pool of ATSBS. As a result of the literature review, the examinations of the special field competencies of science and technology teachers determined by MOE (2008) and the teachers' compositions, the item pool was created with a total of 64 items—45 positive and 19 negative.

2.2.3 Determination of the Measurement Format

Since the classification level and the sensitivity decrease as the number of categories in a scale decreases, and since it becomes more difficult to discriminate as the number of categories increases, it was considered appropriate to prepare the items of the scale in a five-point Likert format. In the answers to the items in the ATSBS, one point was given to answers "I don't agree at all", two points were given to answers "I rarely agree", three points were given to answers "I sometimes agree", four points were given to answers "I mostly agree", and five points were given to answers "I completely agree" for positive items. However, the scale consisted of both positive and negative items to ensure the consistency of the answers given by the participants regarding the construct, and to prevent the partiality for giving positive answers to items. The participants' answers to the items were entered into the SPSS program after the pilot application, and the negative items were transformed via reverse coding, i.e., (1→5), (2→4), (3→3), (4→2), (5→1).

2.2.4 Review of the Pilot Scale by Experts

Before the pilot application of the scale, the draft of the ATSBS was examined by six faculty members of the science teaching department. Firstly, an "expert opinion form" developed within the context of the research was used to elicit opinions and suggestions about whether the draft scale enclosed the theoretical construct. This form was developed to include parts enabling the experts to make their assessments according to Lawshe's (1975) content validity ratio (CVR) and Polit, Beck, & Owen's (2007) kappa coefficient of agreement formula. In the first part of the form, the experts were asked to assess each item of the draft scale as "necessary" "sufficient but should be corrected" or "unnecessary". According to these criteria, each item's content validity index (CVI), probability of chance (Pc) and kappa coefficient of agreement (k) values were calculated using Lawshe's (1975) method for the content validity ratio (CVR) and Polit, Beck and Owen's (2007) kappa statistic formula. It was decided that the items which had zero or negative values of CVR calculated from the assessments of the experts (Yurdagül, 2005) and which had a kappa coefficient of agreement between 0.00 and 0.20, should be excluded from the scale (Fleiss, 1981).

In the second part of the form, the experts were asked to give their suggestions about items which should not be excluded but should be corrected. As a result, a draft scale consisting of a total of 41 items, (29 positive and 12 negative), was formed for the pilot application. Following this, the items were checked by a faculty member of the Turkish teaching department in terms of being simple, understandable and grammatically correct. In line with the opinions of the language expert, some of the items in the draft scale were edited.

2.2.5 Pilot Scale

Before the pilot scale was finalized, a pre-pilot scale was used with 10 science teachers other than those in the research group, in order to identify items which were difficult to understand and to measure the average time required to complete the whole scale. According to the feedback from these teachers, it was found that the scale was not difficult to understand, and the period of time required for answering the scale was about 30-35 minutes.

Finally, participant instructions and a personal information form were added to the scale, forming the final pilot scale.

For factor analysis, the pilot scale was applied to 113 volunteer science teachers working in the province and towns of Ordu during the 2016-2017 academic year. After the scale was implemented, the scale forms which included random answers, unanswered items or multiple answers, were not entered into the SPSS program and the analyses were conducted on the data from a total of 106 participants.

2.2.6 Validity, Reliability and Item Analysis Studies

Reliability analysis was conducted on the ATSBS as a whole scale and on its sub-dimensions by calculating the Cronbach's alpha reliability coefficients. Construct validity was examined with content validity and factor analyses by taking experts' views. In addition, item analysis was conducted with item-total correlation and with an independent groups t-test, in order to determine whether there was a difference between the scores of the items in the lower and upper groups of 27%.

After the exploratory factor analysis (EFA) of the scale was conducted with the help of the SPSS 22.0 program, confirmatory factor analysis (CFA) was conducted using the LISREL 8.51 program. The EFA and the CFA of the scale were both undertaken with the same research group data.

3. Results

3.1 Results of Factor Analysis

Construct validity gives proof about how well the scale measures the concept (factor) which is intended to be measured (DeVellis, 2014). In this study, two kinds of factor analyses were applied to obtain proof about the construct validity of the ATSBS. First, the factor structure of the developed scale was identified, and then CFA was used to test whether the EFA and the determined item-factor structure were consistent.

The varimax technique was applied in EFA (Can, 2014). In order to test the suitability of the data structure in EFA, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity results were examined (Table 1).

Table 1. The results of KMO and Bartlett's test of sphericity

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.79
Bartlett's Test of Sphericity	Approx. chi-Square	615.10
	df	78
	Sig.	0.000

A KMO value of between 0.5 and 0.7 shows that the sample size is sufficient (Can, 2014). Table 1 shows that the KMO sample sufficiency level was 0.79 and Bartlett's test resulted in a chi-squared value ($x^2_{(78)}$ = 615.10; p < 0.001) that was statistically significant. These results show that the sample size of the data is suitable for factor analysis and that the data originate from a multivariate normal distribution (Kaiser, 1974). In conclusion, since these two examined premises are met, it means that there is a suitable quantity of data and thus they are suitable for EFA (Tabachnick & Fidell, 2007).

Principal component analysis (PCA), (a factorization method) and varimax vertical rotation techniques, (a rotation technique), were used to show the factors under which the items were grouped. Following the rotation, the distributions of the items to the factors, factor loads, common variations, averages and standard deviation values are given in Table 2.

Table 2. The results for the factor structure of ATSBS after varimax rotation

Item Direction	Item No.	Factor Loadings			Extraction
		1st Factor	2nd Factor	3rd Factor	
+	32	0.91			0.84
+	37	0.83			0.78
+	33	0.79			0.79
+	36	0.78	0.34		0.78
+	39	0.74			0.68
+	17		0.83		0.71
+	12		0.79		0.68
+	19		0.75		0.68
+	16	0.36	0.74		0.70
-	30			0.84	0.76
-	15			0.84	0.70
-	3			0.78	0.64
-	1			0.77	0.62
Eigenvalues		3.56	2.83	2.79	Total
% of Variance		27.40%	21.76%	21.44%	% of Variance: 70.60%
Item No.		5	4	4	Total Item No.: 13

Note. Suppress absolute values less than 0.32.

Following the varimax rotation, the items in the scale were analysed in terms of the criteria of having a factor load value of 0.32 or lower and being an overlapping item (an item having load value differences of less than 0.10 under more than one factor) (Çokluk et al., 2016). In the pilot form of the ATSBS, 28 items (items 2, 4-11, 13, 14, 18, 20-29, 31, 34, 35, 38, 40 and 41) which were found to have these two properties were excluded from the scale in the light of expert opinions. As a result of excluding these items from the analysis, it was found that the remaining 13 items were grouped into three factors with eigenvalues greater than unity, and that the total contribution of the determined factors to the variance was 70.60%. The factors of a scale are expected to explain at least 52% of the total variance (Henson and Roberts, 2006), and according to this result, it can be said that there is a high level of explained variance. In addition, in terms of factor load, it was found that items in the first factor had values between 0.74 and 0.91, items in the second factor had values between 0.74 and 0.83, and items in the third factor had values between 0.77 and 0.84. According to this result, it can be said that the items in the scale were all highly associated with the relevant factor and the scale had a strong construct. It was found that the common variance values of the items in the ATSBS were between 0.62 and 0.84. If an item's common factor variance is close to unity, this shows that the item's contribution to the variance is high, while if it is close to zero it shows that the contribution is low (Çokluk et al., 2016). Therefore, in this case, the contribution of the items in the scale to the variance was high.

At the stage of assigning factors or deciding on the number of factors, more than one technique can be used. Scree plot analysis, which is one possible technique, highlights important ideas about the factor construct of the scale (Stevens, 2002). When the scree plot is examined, each gap between two points represents a factor, and factors which decrease rapidly are associated with important factor numbers (Field, 2009). The scree plot shows that the scale has three factors (Figure1).
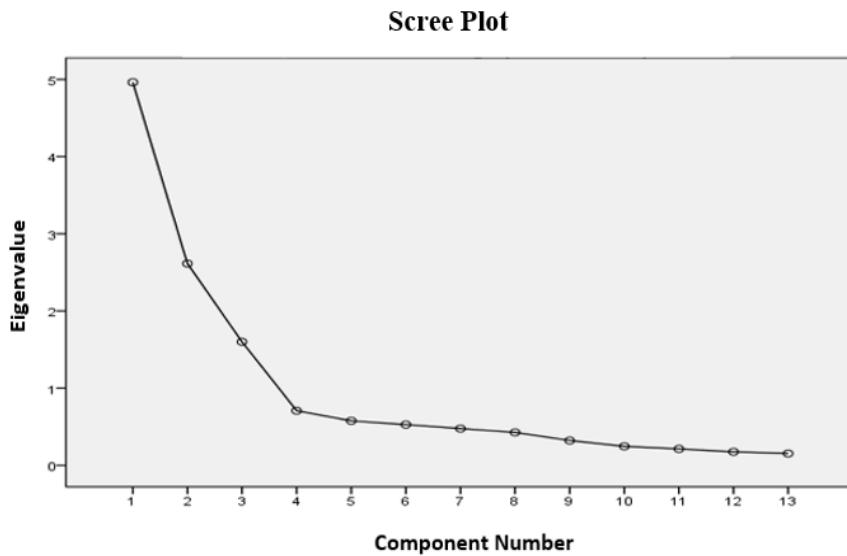
Scree Plot


Figure 1. The scree plot of ATSBS after varimax rotation

After EFA was implemented, CFA was implemented to examine whether the determined ATSBS confirmed the factor construct. In the CFA analysis conducted with the LISREL 8.51 program, modifications were made to some of the items in the scale (17-32 and 32-33). Following the modifications, the fit indices of the scale which were found to have a total of three factors and 13 items are given in Table 3, while path diagram is shown in Figure 2.

Table 3. Model fit indices obtained from the CFA result

χ^2	df	χ^2/df	Sig.	GFI	AGFI	CFI	NNFI	IFI	SRMR	RMSEA
97.00	72	1.35	0.000	0.86	0.79	0.94	0.92	0.94	0.08	0.06

In this study, the chi-squared value obtained as a result of CFA ($\chi^2 = 97.00$) is seen to contribute significantly ($p < 0.001$) (Table 3). It can be said that the chi-squared value and degrees-of-freedom ratio (χ^2/sd) show perfect fit when their values are less than 2, CFI, NNFI and IFI values show perfect fit at 0.95 and above, GFI and AGFI values show perfect fit at 0.90 and above and SRMR and RMSEA values show perfect fit at 0.50 and below (Jöreskog & Sörbom, 1993; Hu & Bentler, 1999; Tabachnick & Fidell, 2007; Kline, 2011). In addition, in the light of the literature (Anderson & Gerbing, 1984; Marsh, Balla & McDonald, 1988), it can be said that the chi-squared value and the degrees-of-freedom ratio (χ^2/sd) show an acceptable fit at values less than 5, CFI, NNFI and IFI values show acceptable fit at 0.90 and above, GFI values show acceptable fit at 0.85 and above, AGFI values show acceptable fit at 0.80 and above and SRMR and RMSEA values show acceptable fit at 0.80 and below.

The chi-squared value and degrees-of-freedom ratio (2/sd = 1.67) obtained following the modification after CFA showed perfect (≤ 2) fit (Anderson & Gerbing, 1984; Marsh et al., 1988). When the other fit indices were examined, it was found that GFI = 0.86, CFI = 0.94, NNFI = 0.92, IFI = 0.94 (≥ 0.90) and RMR = 0.08 and RMSEA = 0.06 (≤ 0.08), all representing good values, while the AGFI (0.79) value was not acceptable (≥ 0.85). One of the reasons why the AGFI value was not acceptable may be that this fit index is quite sensitive to sample size, and may show better fit values with bigger samples (Tabachnick & Fidell, 2007). In conclusion, it can be said that the factor construct of the scale developed as a result of the EFA had an acceptable fit with the data from the CFA.
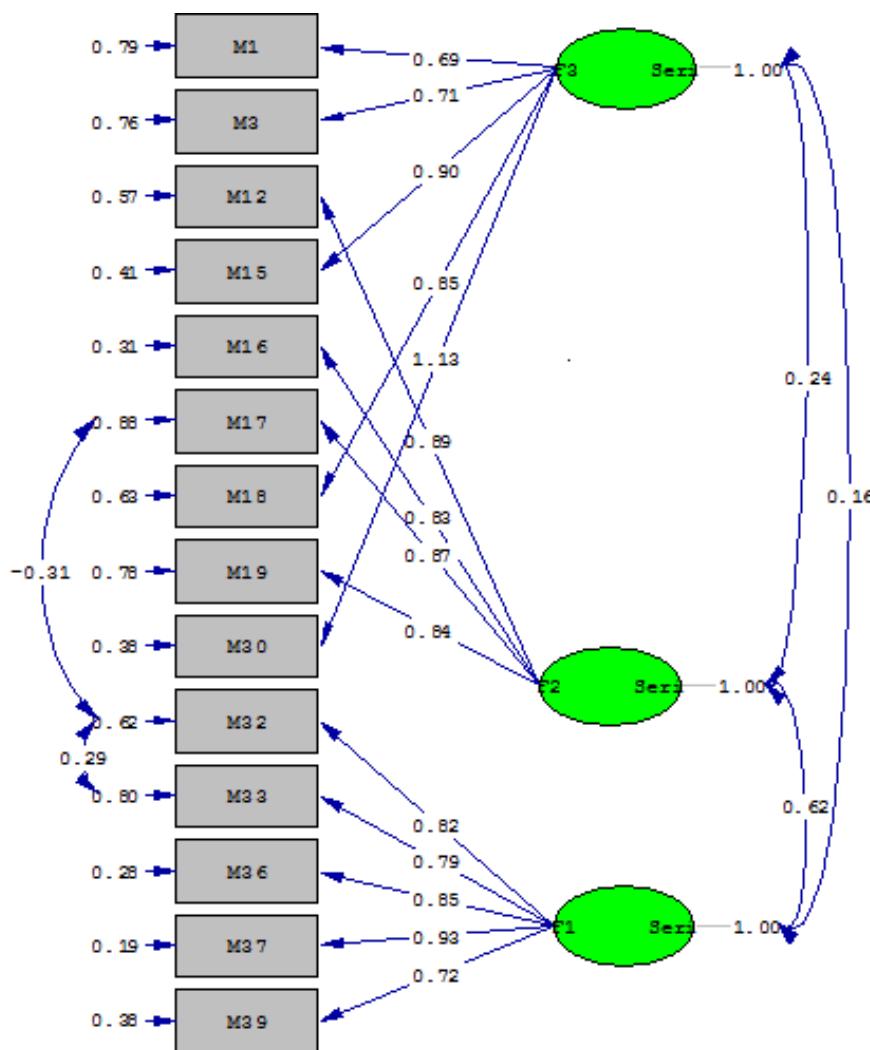
Chi-Square=97.00, df=72, P-value=0.02644, RMSEA=0.064

Figure 2. The ATSBS path diagram after modification

3.2 Results of Item Analysis

In this study, for item analysis, it was examined both item-total correlation and the difference between the item average scores of the 27% upper and lower groups. Within this context, it were analysed through independent groups t-tests the difference between the item average scores of the upper and lower groups with a ratio of 27% for item discrimination analysis (Kelley, 1939). The ATSBS item analysis results for the independent groups t-tests based on item-total correlation and the difference between the 27% lower and upper groups item average scores, are given in Table 4.

Table 4. Item-total correlations and comparison of 27% upper and lower groups

Factors	Mean	Std. Deviation	Item No.	Item-Total Correlation $(r_j)^a$	Cronbach's Alpha if Item Deleted	Mean Score		t (27% Sub-Upper Group)[b]
						27% Lower Group	27% Upper Group	
1st Factor (α=0.90)	3.25	1.10	**32**	0.47[(**)]	0.84	2.70	4.04	4.30[***]
	3.14	1.20	**33**	0.50[(**)]	0.84	2.30	3.87	4.42[***]
	3.15	1.01	**36**	0.72[(**)]	0.84	2.30	4.04	6.73[***]
	3.19	1.04	**37**	0.70[(**)]	0.84	2.52	4.17	6.46[***]
	3.27	0.93	**39**	0.64[(**)]	0.84	2.48	4.04	6.16[***]
2nd Factor (α=0.83)	1.93	1.13	**12**	0.43[(**)]	0.82	1.09	2.52	5.41[***]
	2.27	0.97	**16**	0.65[(**)]	0.83	1.30	3.00	7.98[***]
	2.20	1.34	**17**	0.48[(**)]	0.83	1.13	2.96	6.53[***]
	2.79	1.17	**19**	0.55[(**)]	0.83	1.52	3.70	8.84[***]
3rd Factor (α=0.83)	3.44	1.16	**1**	0.33[(**)]	0.83	2.91	4.26	4.46[***]
	3.31	1.06	**3**	0.31[(**)]	0.82	2.30	3.83	5.40[***]
	3.49	1.10	**15**	0.39[(**)]	0.82	3.04	4.35	4.75[***]
	3.47	1.25	**30**	0.37[(**)]	0.82	2.74	4.39	5.26[***]

Note. [a] n = 106; [b] $n_1 = n_2$ = 29; [**] $p < 0.01$; [***] $p < 0.001$.

An item-total correlation coefficient of 0.30 and above means that items exemplify similar behaviours and the internal consistency of the scale is high (Field, 2009). Table 4 shows that the Pearson correlation coefficients (r) used in the calculation of the item-total correlation of the scale were between 0.31 and 0.72 and they were found to be statistically significant (p< 0.01). In addition, the difference between the item average scores of the 27% upper and lower groups was found to be in favour of the upper group, and was statistically significant (p < 0.001). Therefore, it can be said that the distinctiveness of the items in the ATSBS was good and that these items could be included in the scale as they stood (Atılgan et al., 2015).

Table 5. Averages of ATSBS and sub-factors, standard deviations, correlations and 27% lower and upper group comparisons

Factors	Mean	Std. Deviation	Factor-Total Correlation $(r)^a$	Mean Score		t (27% Lower-Upper Group)[b]	Sig.
				27% Lower Group	27% Upper Group		
1st Factor (α=0.90)	16.02	4.44	0.79[(**)]	10.14	21.21	19.86	0.000[***]
2nd Factor (α=0.83)	9.19	3.85	0.77[(**)]	5.10	14.24	25.53	0.000[***]
3rd Factor (α=0.83)	13.28	3.78	0.56[(**)]	8.86	18.10	21.49	0.000[***]
Total (α=0.84)	38.92	8.55	1	29.00	49.72	18.175	0.000[***]

Note. [a] n = 106, [b] $n_1 = n_2$ = 29; [**] $p < 0.01$; [***] $p < 0.001$.

Table 5 gives the correlation of the sub-factors of the ATSBS with the scores taken from the whole scale and the difference between the item average scores of the 27% upper and lower groups. The results show that the Pearson correlation coefficients were 0.79 for the first factor, 0.77 for the second factor and 0.56 for the third factor, and that the associations between the sub-factors and the whole scale were positive, high and significant. In addition, as a result of the 27% lower and upper group comparison of the ATSBS sub-factors, it was found that t-values in all the lower groups differed significantly (p < 0.001) in favour of the upper group. According to all these results, it can be concluded that the scale's sub-factors had high validity, and that they all measured the same construct.

Table 6. Correlations between ATSBS and sub-factors

	1st Factor	2nd Factor	3rd Factor	Total	Item No.
1st Factor	1	0.50[**]	0.11[**]	0.79[**]	5
2nd Factor	0.50[**]	1	0.15[**]	0.77[**]	4
3rd Factor	0.11[**]	0.15[**]	1	0.56[**]	4
Total	0.79[**]	0.77[**]	0.56[**]	1	13

Note. n = 106; [**] $p < 0.01$.

In addition, Cronbach's alpha coefficients were calculated for the whole scale and for each sub-factor, in order to analyse how consistent (in terms of internal consistency) the items of the scale were. In general, it can be said that for Cronbach's alpha reliability coefficient values of 0.70 and above, the scale developed is reliable (Fraenkel et al., 2012). The Cronbach's alpha reliability coefficient values were 0.84 for the whole scale, 0.90 for the first factor, and 0.83 for the second and third factors.

The three sub-factors of the ATSBS obtained as a result of factor analysis and item analysis were named within the theoretical framework by taking into consideration the opinions of two faculty members in science education. Table 7 shows the names of the ATSBS factors and the items in each factor.

Table 7. ATSBS factors and the items in each factor

Factor	Item No	Item
Student Outcomes Through Astronomy Teaching (1st Factor)	32	I can develop students' skills in questioning information about astronomy
	33	I can teach the target behaviours about astronomy to students who are experiencing difficulty in learning.
	36	I am effective in making students gain contemporary and scientific information about astronomy.
	37	I teach realistic and scientific opinions to students.
	39	I develop students' skills in making comments about astronomy subjects.
Astronomy Teaching Strategies (2nd Factor)	12	I can organize extracurricular activities about astronomy for my students.
	16	I can organize classroom experiments or activities about astronomy.
	17	I can teach astronomy subjects by using various virtual-reality programs (Stellarium, Celestia, etc.).
	19	I can teach astronomy subjects by using scientific process skills (using the association between space and time, observation, etc.).
Difficulty in Astronomy Teaching (3rd Factor)	1[*]	I have difficulty in explaining information about astronomy in the light of scientific information.
	3[*]	I have difficulty in commenting on astronomy subjects.
	15[*]	I have difficulty in teaching astronomy concepts by associating them with real life.
	30[*]	I have difficulty in choosing teaching methods and techniques suitable for students' individual differences.

Note. [*] Negative items.

Lastly, it was found that the lowest score obtainable from the 13-item ATSBS was 13, while the highest score was 65. In addition, in the light of experts' opinions, it was found that five minutes could be sufficient time to complete the ATSBS.

4. Conclusion, Discussion and Suggestions

The purpose of this study was to develop a reliable and valid assessment instrument which measures the self-efficacy beliefs of science teachers about the teaching of astronomy subjects. After the application of all the scale development stages, it was found that the scale consisted of a total of 13 items using a five-point Likert format with three factors.

As a result of the pilot scale, a factor analysis was conducted to investigate the construct validity of the ATSBS. First of all, the KMO and Bartlett's test of sphericity were used to test whether the data obtained were suitable for factor analysis. Analysis results showed that the KMO value was 0.79 and Bartlett's test of sphericity was found to be statistically significant (p < 0.001). These results show that the data and the sample size were suitable in terms of the applicability of factor analysis and the data were found to be normally distributed. In

other words, the variables had associations high enough to form a reasonable basis for their use (Leech, Barrett, & Morgan, 2005).

The data tested for suitability for factor analysis were first examined with EFA, and then with CFA, to confirm the factor construct of the scale. Principal components analysis was adapted and varimax (maximum change) rotation was used. As a result of varimax rotation, it was concluded that the ATSBS had a total of 13 items -nine positive and four negative- and three factors. In addition, it was found that the factors of the scale explained 70.60% of the total variance. According to this result, it can be said that the items in the factors were highly correlated with the factor and the scale had a strong construct.

After the scale's factor construct was determined with EFA, CFA was applied. As a result of the findings from the fit indices of the construct with three factors, the ratio (x^2/sd = 1.35) of the chi-squared value (x^2 = 97) to the degree of freedom (sd = 72) was found to be perfect, while the other fit indices (RMSEA = 0.06, SRMR = 0.08, CFI = 0.94, IFI = 0.94, NNFI = 0.92, AGFI = 0.79 and GFI = 0.86) were found to show an acceptable fit, and the tested model was concluded to be sufficient.

Within the reliability and validity analysis context of the scale, the correlations of both the items and the factors of the ATSBS with the total score were examined, together with whether the difference between the average scores of the 27% upper and lower groups was significant. Within this context, it was found that the item-total correlation coefficients of the scale were between 0.31 and 0.72 and the factor-total correlation coefficients were between 0.56 and 0.79. Similarly, the difference between the item average scores of the 27% upper and lower groups was analysed using the independent groups t-test. According to the t-test results for the lower and upper groups, the scale was found to have significant levels (p < 0.001) for the lower and upper groups on the basis of both items and factors. The correlation coefficients of each factor of the scale with the others were found to be positive and significant (p < 0.001). Therefore, it can be said that the items in the ATSBS can differentiate well between individuals and that the items measure similar behaviours. In addition, the Cronbach's alpha reliability coefficients were measured to test the internal consistency of the ATSBS. The Cronbach's alpha values were found to be 0.84 for the whole scale, 0.90 for the first factor and 0.83 for the second and third factors.

Items were grouped under the factors of ATSBS which were found to be reliable and valid. The factors were named after being examined by two faculty members in science education and after consulting similar studies in the literature. In conclusion, the first factor of the scale was named "student outcomes through astronomy teaching", the second factor was named "astronomy teaching strategies" and the third factor was named "difficulty in astronomy teaching".

When the studies in the literature were examined, one study was found which measured self-efficacy beliefs with respect to astronomy instruction (Güneş, 2010). The "Science Instruction Self-Efficacy Belief Scale" which was developed for pre-service teachers by Riggs & Enochs (1990) and adapted into Turkish by Özkan et al. (2002) was revised by Güneş (2010) by changing the word "science" to "astronomy", and the "Astronomy Instruction Self-Efficacy Belief Scale" was developed. Güneş (2010) tested the 23-item scale in terms of reliability and validity by administering it to prospective science and social science teachers. Güneş (2010) found that the Cronbach's alpha reliability coefficients were 0.80 for the whole scale, 0.87 for the astronomy instruction personal self-efficacy factor and 0.78 for the astronomy instruction outcome expectation factor. When this scale was compared with the ATSBS developed for this study in terms of reliability coefficients, it could be seen that they were similar in terms of being highly reliable and that the reliability levels of the whole scales were parallel. In terms of the sub-factors of the scale, the student outcomes through astronomy teaching factor of the ATSBS was partly similar to the astronomy instruction outcome expectations factor in terms of the measurability characteristics of behaviours, and also in terms of the high values of the Cronbach's alpha reliability coefficient in both cases. However, the ATSBS is different from the scale in the literature in that it was developed for science teachers and in that the items were completely original. In conclusion, a reliable and valid assessment tool was developed which measured the self-efficacy beliefs of science teachers with respect to the teaching of astronomy subjects.

As a result of the findings of the study, the following suggestions are made for researchers and those who administer the scale:

- Various studies which include both qualitative and quantitative methods and which aim to show the self-efficacy beliefs of science teachers regarding astronomy subjects could be undertaken. Within this context, the ATSBS could be used as an assessment instrument, and from this and other qualitative data collection instruments, detailed information about the self-efficacy beliefs of teachers could be obtained.

- The correlation between the ATSBS and other assessment instruments which measure field information, pedagogic field information, attitudes and motivation, etc., and thus model how such variables predict the self-efficacy beliefs of teachers about astronomy teaching, could be presented. Active practices for teachers could be implemented with the developed model.

- This scale was developed within the context of a Turkish sample. The scale could be adapted for, and used in, different cultures.

- Since basic astronomy subjects are taught in the third and fourth grades, ATSBS could be adapted for primary school teachers, thus determining their self-efficacy beliefs about the teaching of astronomy subjects.

- Teachers should be able to self-assess, and thus become aware of their own shortcomings. They should be given opportunities to develop themselves in areas where they are weak.

References

Agan, L. (2004). Stellar Ideas: Exploring Students' Understanding of Stars. *Astronomy Education Review, 3*(1), 77-97. https://doi.org/10.3847/AER2004008

Ali-Shah, S. (2009). *Impact of teacher's behavior on the academic achievement of university students.* Doctoral Thesis, University Institute of Education and Research, University of Arid Agriculture, Rawalpindi, Pakistan.

Atılgan, H., Kan, A., & ve Doğan, N. (2015). *Eğitimde ölçme ve değerlendirme* (8. Baskı). Ankara: Anı Yayıncılık.

Azar, A. (2010). Ortaöğretim fen bilimleri ve matematik öğretmeni adaylarının öz yeterlilik inançları. *Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Dergisi, 6*(12), 235-252.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(52), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy: the exercise of control* (p. 604). New York: W. H. Freeman and Company.

Barros, M. A., Laburù, C. E., & da Silva, F. R. (2010). An instrument for measuring self-efficacy beliefs of secondary school physics teachers. *Procedia—Social and Behavioral Sciences, 2*(2), 3129-3133. https://doi.org/10.1016/j.sbspro.2010.03.476

Baxter, J. (1989). Children's understanding of familiar astronomical events. *International Journal of Science Education, 11*, 502-513. https://doi.org/10.1080/0950069890110503

Blonder, R., Benny, N., & Gail-Jones, M. (2014). Teaching self-efficacy of science teachers. In R. Evans, J. Luft, C. Czerniak, & C. Pea (Eds.), *The Role of Science Teachers' Beliefs in International Classrooms* (pp. 3-16). Rotterdam, Boston, Taipei: Sense Publishers. https://doi.org/10.1007/978-94-6209-557-1_1

Çokluk, Ö., Şekercioğlu, G., & ve Büyüköztürk, Ş. (2016). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları.* Ankara: Pegem Akademi.

DeVellis, R. F. (2014). Ölçek Geliştirme: Teori ve Uygulamalar. Çev. Ed.: Totan, T., Ankara: NOBEL Akademik Yayıncılık (Orijinal çalışma basım tarihi 2012).

Diakidoy, N., & Kendeou, P. (2001). Facilitating conceptual change in astronomy: a comparison of the effectiveness of two instructional approaches. *Learning and Instruction, 11,* 1-20. https://doi.org/10.1016/S0959-4752(00)00011-6

Enochs, L. G., & Riggs, I. M. (1990). *Further development of an elementary science teaching efficacy belief instrument: A preservice elementary scale.* Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (63rd, Atlanta, GA, April 8-11, 1990). https://doi.org/10.1111/j.1949-8594.1990.tb12048.x

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd edition). London: SAGE Publications.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378-382. https://doi.org/10.1037/h0031619

Ford, R. (2002). *Teacher self-efficacy and its influence on student motivation ındıa.* Doctoral Thesis, Bachelors of Science in Education, Indiana University, Bloomington.

Fraenkel, J., Wallen, N., & Hyun, H. H. (2012). *How to design and evaluate research in education.* Boston: McGraw Hill.

Friedman, I. A., & Kass, E. (2002). Teacher self-efficacy: a classroom-organization conceptualization. *Teaching and Teacher Education, 18*, 675-686. https://doi.org/10.1016/S0742-051X(02)00027-6

Gavora, P. (2010). Slovak pre-service teacher self-efficacy: theoretical and research considerations. *The New Educational Review, 21*(2), 17-30.

Gibson, S. (1983). *Teacher efficacy: a construct validation study.* Doctoral Thesis, Faculty of the Graduate School, University of Southern California, Losdisse Angeles, California.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: a construct validation. *Journal of Educational Psychology, 76*(4), 669-682. https://doi.org/10.1037/0022-0663.76.4.569

Güneş. G. (2010). Öğretmen adaylarının temel astronomi konularında bilgi seviyeleri ile bilimin doğası ve astronomi özyeterlilikleri arasındaki ilişkinin incelenmesi. Yüksek Lisans Tezi, Çukurova Üniversitesi, Sosyal Bilimler Enstitüsü, İlköğretim Anabilim Dalı, Adana.

Henson, R. K. (2001). Teacher self-efficacy: substantive implications and measurement dilemmas. The Annual Meeting of the Educational Research Exchange, 26 January, 2001, College Station, Texas.

Henson, R. K., & ve Roberts, J. K. (2006). Use of exploratory factor analysis in published research: common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393-416. https://doi.org/10.1177/0013164405282485

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31-36. https://doi.org/10.1007/BF02291575

Kaya, V. H., Polat, D., & Karamüftüoğlu, İ. O. (2014). Development of Self-Efficacy Scale For Science Education. *The Journal of Academic Social Science Studies, 28*, 581-595.

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*(1), 17-24. https://doi.org/10.1037/h0057123

Kiremit, H. Ö. (2006). Fen Bilgisi Öğretmenliği Öğrencilerinin Biyoloji İle İlgili Öz Yeterlik İnançlarının Karşılaştırılması. Yayınlanmamış Doktora Tezi. İzmir: Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28,* 563-575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for intermediate statistics: Use and interpretation* (2nd edition). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Midgley, C., Feldlaufer, H., & Eccles, J. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81,* 247-258. https://doi.org/10.1037/0022-0663.81.2.247

MOE. (2008). Fen ve teknoloji öğretmeni özel alan yeterlikleri: Öğretmen yeterlikleri öğretmenlik mesleği genel ve özel alan yeterlikleri. Ankara: Devlet Kitapları Müdürlüğü.

Peña, B. M., & Quílez, M. J. (2001). The importance of images in astronomy education. *International Journal of Science Ecucation, 23*(11), 1125-1135. https://doi.org/10.1080/09500690110038611

Plummer, J., & Zahm, V. (2010). Covering the Standards: Astronomy Teachers' Preparation and Beliefs. *Astronomy Education Review, 9*(1), 1-28. https://doi.org/10.3847/AER2009077

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Focus on research methods is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*, 459-467. https://doi.org/10.1002/nur.20199

Protheroe, N. (2008). Teacher efficacy: what is it and does it matter? *Principal, 87*(5), 42-45.

Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Teacher Education, 74*(6), 625-637. https://doi.org/10.1002/sce.3730740605

Ritter, J. M., Boone, W. J., & Rubba, P. A. (2001). Development of an instrument to assess prospective elementary teacher self-efficacy beliefs about equitable science teaching and learning (SEBEST). *Journal of Science Teacher Education, 12*(3), 175-198. https://doi.org/10.1023/A:1016747713585

Roberts, J. K., & Henson, R. K. (2000). Self-efficacy teaching and knowledge instrument for science teachers (SETAKIST): A proposal for a new efficacy instrument. Paper presented at the Mid-South Educational research Association (28th, Bowling Green, KY, November 17-19, 2000).

Rodger, S., Murray, H. G., & Cummings, A. L. (2007). Effects of teacher clarity and student anxiety on student outcomes. *Teaching in Higher Education, 12*(1), 91- 104. https://doi.org/10.1080/13562510601102255

Ross, J., & Bruce, C. (2007). Professional development effects on teacher efficacy: Results of randomized field trial. *Journal of Educational Research, 101*(1), 50-60. https://doi.org/10.3200/JOER.101.1.50-60

Seçer, İ. (2015). *Psikolojik test geliştirme ve uygulama süreci.* Ankara: Anı Yayıncılık.

Show-Alter, B. S. (2005). *The effect of middle school teachers' mathematics teaching self-efficacy beliefs on their students' attitudes toward mathematics.* Doctoral Thesis, Faculty of the Faculty of the Graduate College, University of Oklahoma, Norman, Oklahoma.

Sia, A. (1992). Preservice elementary teachers' perceived efficacy in teaching environmental education: A preliminary study. Paper presented at the North American Association for Environmental Education Annual Conference in Toronto, October 20, Canada.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th edition). Hillsdale, NJ: Erlbaum.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5th edition) New York: Pearson Education, Inc.

Tschannen-Moran, M., & Woolfolk-Hoy, A. (2001). Teacher efficacy: capturing an elusive concept. *Teaching and Teacher Education, 17*, 783-805. https://doi.org/10.1016/S0742-051X(01)00036-1

Tschannen-Moran, M., Woolfolk-Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202-248. https://doi.org/10.3102/00346543068002202

Yurdugül, H. (2005). Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması. XIV. Ulusal Eğitim Bilimleri Kongre Kitabı, Denizli, 1-6.

Zeilik, M., Schau, C., Mattern, N., Hall, S., Teague, K. W., Bisard, W. (1997). Conceptual astronomy: a novel model for teaching post secondary *science courses. American Journal of Physics, 65*(10), 986-996. https://doi.org/10.1119/1.18702

Copyrights