

Developing a Scale to Measure Students' Attitudes toward Science

Adem Akkuş ^{1,*}

¹ Mus Alparslan University, Education Faculty, Elementary Science Education, Mus, Turkey

ARTICLE HISTORY

Received: 03 April 2019

Revised: 31 October 2019

Accepted: 06 December 2019

KEYWORDS

Science Attitude,
Scale Development,
Scale validation

Abstract: The aim of this study is to develop a science attitude scale (SAS). For that purpose, the literature review has been done for suggestions for creating scales and a new draft scale developed. The draft scale was analyzed by specialists and a pilot study is done after its approval by experts. The SAS is prepared with 21 items and among these, 11 items are reverse-coded. The SAS consists of Likert-type items. The sample of the study consists of 154 college students studying at the Faculty of Education, Elementary Science Education, and Elementary Education departments. Principal axis factoring with orthogonal rotation (varimax) was used for exploratory factor analysis. Factor eigenvalues were checked with respect to parallel analysis and numbers of the factors were determined with respect to the analysis. Items that did not serve the purpose of the scale were omitted from the SAS. The finalized SAS' Cronbach alpha value is .953. For confirmatory factor analysis data were collected from a different sample which consists of university students who were studying at elementary science education, elementary education, and electric electronic engineering departments. Number of sample is 201. Confirmatory factor analyses run through Amos 24.0 software. It is believed that SAS is a valuable contribution to the science education field since it has unidimensional structure and proved its item discrimination power, and alongside with an excellent internal consistency. SAS also offers opportunity to develop multidimensional science attitude scale. For that purpose, original SAS and English version of it are provided in appendixes.

1. INTRODUCTION

Attitude is defined as an individual's positive or negative characteristics towards a subject (Serin & Mohammadzadeh, 2008). Students with positive attitudes toward science are likely to display more science-related attitudes and choose science-related professions. On the other hand, recent studies indicate that there is a trend in that science-related departments attract fewer students than social science-related departments (Shah & Mahmood, 2011). Therefore, attitudes toward science and science-related subject areas are in focus of research studies. Even science attitudes may be used for predicting science achievement (Adesoji, 2008). Factors affecting attitudes are also among the subjects to be studied. For example, gender might be suggested as one of the factors. Although both genders have closely similar attitude values

CONTACT: Adem AKKUŞ ✉ ademakkus@gmail.com 📧 Mus Alparslan University, Education Faculty, Elementary Science Education, Mus, TURKEY

ISSN-e: 2148-7456 /© IJATE 2019

toward science, underlying factors might be different. Girls learn better in an organized environment and boys' attitudes are related to cohesiveness. Other factors might be listed as instructional style, teaching strategies, classroom design, etc. (Bernardez, 1982). Thus, having knowledge of students' attitudes and encouraging them toward science is important and attitudes of students must be known (Shah & Mahmood, 2011). Scales are useful for this purpose and in this regard, researchers try to develop their own instruments for various purposes or use a standardized version (Coll, Dalgety & Salter, 2002). Using standardized scales or other means of standardized measurements could value the purpose, letting researchers have an opinion on the attitudes of students and understand dimensions and their value within the context (Demirbaş, 2009). On the other hand, standardized scales are mostly in English and have different theoretical aspects, different cultural settings, psychometric properties and hence may lack assess the right domain of interest, not be suitable for local use due to contextual differences (Shah & Mahmood, 2011). Perhaps that is the reason why different researchers have failed to confirm the Test of Science Related Attitudes scale (TOSRA) in their sample population. Attitudes may be observed in different types of responses and even be affected by curriculum changes (Cheung, 2007). As curriculum changes are made, the need for to measure attitudes and to develop new scales also becomes a value of interest to observe the effect of the curriculum. Even instructional techniques might affect students' attitudes whose change could value the future implications (Evrekli, İnel, Balım & Kesercioğlu, 2009). Since Turkey has already announced that changes on the curriculum are done to promote active learning (TTK, 2017), it is important to observe the effects of curriculum changes on students' attitudes. For that aim, several researchers already tried to develop attitude scales or applied existing ones. For example, Can and Şahin (2015) studied kindergarten teacher candidates' attitudes toward science and science teaching. Analyses were done to investigate the relationships of grades and gender with science attitude and science teaching. Serin & Mohammadzadeh (2008) used a scale to determine attitude and academic success relationships. Korkmaz, Şahin and Yeşil (2011) tried to investigate attitude toward scientific research. For that reason, they developed a scale with 30 items and four dimensions. Tortop (2013) adapted a scale into Turkish for assessing scientific field trip attitude. The study reveals that a single attitude might have different dimensions. For example, another study tried to investigate the relationships between attitudes and science process skills (Dönmez & Azizoğlu, 2010). All these studies show that scales might be used for collecting data (Deshpande, 2004; Hinkin, 1998; Wong & Lian, 2003; Francis et. al., 2004) so that effective measures might be taken into account for this purpose (Hinkin, 1998; Hinkin, Tracey, Enz, 1997). Thus, the purpose of this study is to develop a science attitude scale (SAS). Attitudes may have different dimensions and scales may reflect those dimensions. However, most scales determined the number of dimensions based on eigenvalues through factor analysis. SAS also determined number of dimensions through parallel analysis which reflects more accurate number of dimensions. Moreover, SAS is a unidimensional scale but offers the chance for researchers to develop multidimensional scales based on SAS.

2. METHOD

2.1. Research Design

In the method, to develop a scale, based on suggestions from the literature, the guidelines have been determined (Brinkman, 2009; Deshpande, 2004; Hinkin, 1998; Hinkin et. al., 1997; Johanson & Brooks, 2010; Ajzen, 2005b; Francis et. al., 2004; Cabrera-Nguyen, 2010; Hof, 2012). Those guidelines are:

- a) Not to cause any bias, the items' context must be within the students' cultures (schemes).
- b) Respondents should place themselves at a position.

- c) Items must assess a single behavior or response.
- d) Items must not be interpreted in different ways.
- e) Language and expected knowledge should be familiar with the target group.
- f) Sensitive and double negative items should be avoided.

Among Thurstone's method of equal-appearing intervals, Likert scale, semantic differential scales; it is determined that a Likert type scale would be more beneficial for the purpose of the study to ensure easy compilation and generalization (Lovelace & Brickman, 2013; Brinkman, 2009; Johanson & Brooks, 2010; Hof, 2012). To ensure content adequacy and avoid fatigue, a maximum number of items is determined so that respondents will respond within attention time. For this purpose, the sentences "Strongly agree" or "Strongly disagree" are given at the beginning of the scale as information. By placing five levels of response for an item, it is ensured that internal consistency is increased and sufficient variances are obtained (Hinkin, 1998; Lovelace & Brickman, 2013; Brinkman, 2009; Hinkin et al., 1997; Ajzen, 2005b; Francis et al., 2004). Since there might be respondents tending to choose options at the edges or in the same direction, reversed coded sentences are appropriately used to trigger their vigilance (Hinkin, 1998; Francis et al., 2004; Hof, 2012)

The process of developing the science attitude scale (SAS): The item sentences were finalized after determining SAS' scope, content, items and their numbers. After that specialist views were taken account. The draft science attitude scale (SAS) consisted of 31 items. However, items 25 and 27 were removed from the SAS since they were the same as items 6 and 5. Initial internal reliability analysis was carried out and Cronbach's α value was found as .861 (good according to Kalaycı, 2010). The draft scale's content and scope were analyzed by instructors who have the experience of teaching and researches on related issues since specialists could value the prepared scale on the content domain. Specialists work in the education faculty at science education department (Hinkin et al., 1997). Specialists' views' on sentences and corrected item-total correlation values of the items were cross-checked, and the items regarded as problematic were excluded from the scale. Item of 1 and 9 contradicted guideline f "Sensitive and double negative items should be avoided", items of 8, 11, 14, and 26 contradicted guideline c "Items must assess a single behavior or response" and guideline d "Items must not be interpreted in different ways". Thus, those items were excluded from the SAS immediately for further analysis; thus finalized SAS Cronbach's α reliability value is .943 and with 23 items. The developed SAS consists of twelve reversed questions (items) which are items 2, 5, 6, 7, 10, 16, 17, 20, 21, 22, 23, and 31.

2.2. The sample size and sampling method

The SAS was applied to 154 college students at the Faculty of Education, Elementary Science Education, and Elementary Education departments. In order to ensure the privacy of personal information, (i.e. avoiding conflict of interest) only the students' gender and age information were demanded.

In literature, to determine a sample size has been a controversial issue. Some researchers argue about arbitrary sampling which presents high communalities without cross-loadings. Thus, sampling may be determined by the nature of the data i.e. More acceptable view, some researchers claim that if data is strong enough then sample size might be small, while others argue on item-ratio sampling. The debate on item-ratio suggests proportion from 1:2 to 1:10 for item and sampling (Anthoine, Moret, Regnault, Sébille & Hardouin, 2014; Hinkin, 1998; Hinkin et al., 1997; Cabrera-Nguyen, 2010). Since the item respondents' ratio of the study is 1:7, it is believed that sampling is adequate for the study with respect to first view.

As for arbitrary sampling, several arguments might be stated. For example, Johanson & Brooks (2010) suggests to social researchers that minimum participants are 100 people for sampling.

For a comprehensive item analysis, sample size 100 to 200 person should be conducted since standard errors for Cronbach's alpha value increase as the sample size decreases. However, it is also noted that regardless of the number of items, the mean inter-item correlation is nominal between $N=30$ and 200. Hinkin (1998) and Hinkin et al., (1997) suggest $N=150$ to obtain sufficient data for exploratory factor analysis as long as item inter correlations are reasonably strong. However, researchers also note the difference between statistical and practical significance because attaining statistical difference chance increases as the sample size increases. Larger samples are useful to detect small fluctuations. On the other hand, as sample size increases the practical meaning of the results may distort, so the decision on sample size must be made with caution. Francis et al. (2004) state that some researchers claim that $N=25$ would be enough for purposive sampling and sample size could be increased until it is believed that data saturation has been achieved. Yet, researchers claim a sample size $N=80$ would generally be enough. Cabrera-Nguyen (2010), while stating similar statements, also indicates that some researchers argue on sampling size and claim sample size depends on the gathered data, and adequacy of sampling is determined after analyzing the gathered data. In the same paper, the researcher embraces a mixed approach based on communalities value. Hof (2012) suggests 10-15 respondents per item, yet states a KMO value already signals whether the sample size is enough or not. Based on the suggestions above, it is again regarded that the sample of the study will be enough for the purpose.

3. RESULT

3.1. Reliability Analysis

The data analyzed with respect to internal consistency, communalities, and factor loadings. Analyses were carried out together for better judgment of retaining factors. Corrected item-total correlation values of the items were analyzed. It was observed that items 12 (.004) and 16's (.126) corrected item-total correlation values were below the desired value of .200 (Johnson & Morgan, 2016), hence; these items were excluded from the scale. After this process the scale's Cronbach's α value was found as .953 and "excellent" for the final version of SAS.

3.2. Exploratory Analysis

Exploratory Factor Analysis: A principal axis was conducted on the 21 items with orthogonal rotation (varimax) through the SPSS program to reveal the factors within the developed scale since it is suggested for more reliable scale evaluation (Field, 2013; Hof, 2012). The Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy as "marvelous" (Kalaycı, 2010). The KMO value is .951 and above the acceptable limit of .5 (Field, 2013). Bartlett's test of sphericity was found significant ($X^2(210) = 2392.067$, $p = .00 < .05$). Hence, the KMO value already signaled that the sample size might be enough, the analysis of each SAS item was initiated. An initial analysis was run to obtain eigenvalues for each factor in the data. Three factors emerged having eigenvalue over Kaiser's criterion of 1 and in combination explained %58,778 of the variance. Eigenvalue of the factors were 11,301; 1,340 and 1,017 respectively for factor 1, factor 2 and factor 3.

The parallel analysis offers a good interpretation of the number of retaining factors (Field, 2013; Johnson & Morgan, 2016) thus; a Monte Carlo PCA for parallel analysis with 1000 replications was run to confirm the eigenvalues (Watkins, 2000). Eigenvalues obtained were 1,7259; 1,5915 and 1,4902 for that reason, it was concluded that only the first eigenvalue was acceptable since the second and third factor's eigenvalue was not significant. Therefore, an EFA was rerun with a one-factor solution. The variance shared by the factor was 51,937. The scree plot (Figure 1) was obtained and it was decided that the scale has one factor with respect to the convergence

of the scree plot and Kaiser's criterion on this value. Table 1 shows the factor loadings after rotation and extracted communalities.

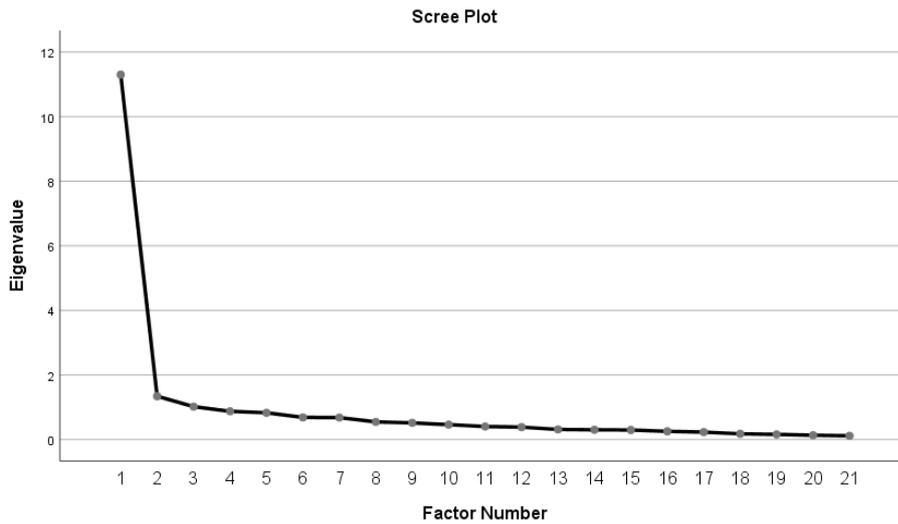


Figure 1. Scree Plot

Table 1. extracted communalities and factor loadings

Item	h^2	Factor
SAS2	,215	,464
SAS3	,602	,776
SAS4	,568	,754
SAS5	,422	,650
SAS6	,474	,689
SAS7	,259	,509
SAS10	,174	,417
SAS13	,494	,703
SAS15	,787	,887
SAS17	,621	,788
SAS18	,295	,543
SAS19	,826	,909
SAS20	,773	,879
SAS21	,549	,741
SAS22	,671	,819
SAS23	,721	,849
SAS24	,624	,790
SAS28	,207	,455
SAS29	,817	,904
SAS30	,617	,785
SAS31	,190	,435

Since none of the items' factor loadings were below .400 (Table 1) validation of scale's internal consistency reliability coefficient was made as suggested (Field, 2013; Francis et. al., 2004). Cronbach's α value was found as =.953 and it means "excellent" (Kalaycı, 2010). For a detailed analysis of items on discrimination value, an independent samples t-test was run for each item. Lower and upper %27 of the samples (N=42) were compared through independent samples t-test. This analysis shows items' discrimination value of individuals between lower and

upper %27 of the sample and is used by many scientists in scale developments (Moore & Foy, 1997). Reliability analysis of item-total correlation and $t_{\text{up-down}(\%27)}$ results and items' codes are shown in Table 2.

Table 2. Item-total correlation and $t_{\text{up-down}(\%27)}$ results

Item	Mean	Standard Deviation	Corrected-item total correlation	$t_{\text{up-down}(\%27)}$
SAS1	2,92	1,603	,456	5,767*
SAS2	3,14	1,500	,757	12,781*
SAS3	3,05	1,522	,738	12,796*
SAS4	3,14	1,650	,643	9,729*
SAS5	3,18	1,602	,677	14,399*
SAS6	3,23	1,475	,506	6,588*
SAS7	3,06	1,423	,412	4,944*
SAS8	3,07	1,417	,680	9,175*
SAS9	3,31	1,619	,863	26,213*
SAS10	3,20	1,479	,765	12,029*
SAS11	3,06	1,538	,520	7,381*
SAS12	3,19	1,555	,889	22,715*
SAS13	3,38	1,500	,862	23,059*
SAS14	3,03	1,434	,722	13,191*
SAS15	3,17	1,564	,801	16,176*
SAS16	3,26	1,385	,829	17,507*
SAS17	3,22	1,515	,771	13,456*
SAS18	3,16	1,380	,444	5,343*
SAS19	3,19	1,443	,887	20,686*
SAS20	3,08	1,412	,764	14,565*
SAS21	3,01	1,186	,414	6,093*

* p < .05

The finalized SAS consists of 21 items and 11 items are reverse coded items. The reversed coded items are 1, 4, 5, 6, 7, 10, 13, 14, 15, 16 and 21 (Table A1). For international readers, an English translation of SAS is given in Table A2. Translation was done by the researcher and to ensure translation was done correctly and comprehension of the scale is easy, SAS was presented to a professor to take account of specialist's opinion. That professor was working at university and had a formal education in English language. After that a retranslation and crosscheck were done by another professor who also had a formal education in English language and working at Education faculty.

3.3. Confirmatory Factor Analysis

For confirmatory factor analysis data were collected from a different sample. The sample consists of university students who were studying at elementary science education, elementary education, and electric electronic engineering departments. The number of sample is 201 in total. Confirmatory factor analyses run through Amos 24.0 software. Initial analysis results revealed that χ^2/DF ratio is 2,635 RMSEA value is .09; GFI value is .802; CFI value is .627; SRMR value is .0852; NFI value is .520; AGFI value is .758. RMSEA, GFI, AGFI values showed model did not show a good fit with respect to indices values. However, it is noteworthy that they are close to the desired value. CFI and NFI values did not fall into the categories of well fit. Thus, in the review of the literature it is decided to examine the error terms and decide whether some items correlate together or not for a better comprehension of the model. Decisions to covariate items based on the rule of thumbs. First, covariated items should exhibit similar set of idea/pattern/mindset in phrases. Second, number of covariated items should be restricted up

to 7. Third, each time a covariate is done then, the results would be checked to see if model fit indices were changed dramatically or not. Fourth, covariances between the items should not be lower than 10. Covariated items are 13 and 14; 9 and 11; 12 and 17; 18 and 19. Final data analyses results revealed that χ^2/DF ratio is 1.893; RMSEA value is .067; SRMR value is .0746; IFI value is 0.806; GFI value is .852; CFI value is .800; NFI value is .662; PNFI value is 0.583; NNFI (TLI) value is 0.773; AGFI value is .815. The confirmatory factor analysis result is shown in Figure 2.

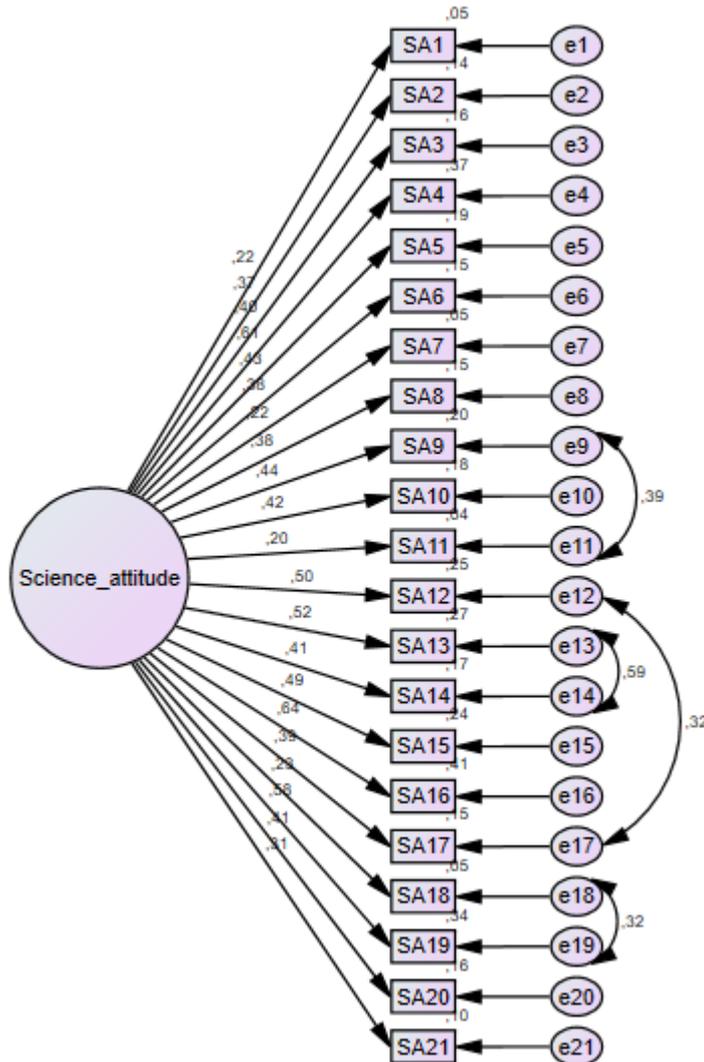


Figure 2. Confirmatory Analysis Result

χ^2/DF ratio is 1.893 and it is regarded that a model has a good fit if Chi-square (χ^2)/degree of freedom (df) ratio is < 2 . RMSEA value is .067 and model has a good fit since $RMSEA \leq 0.1$. IFI value is 0.806 and CFI value is .800 and it is accepted model has a good fit since $CFI \geq 0.8$ and $IFI \geq 0.8$. (Browne and Cudeck, 1993; Garson, 2006 as cited in Chinda, Techapreechawong & Teeraprasert, 2012).

Many recommendations are being done on value of Root Mean Square Error of Approximation (RMSEA). For example, Pedroso et. al. (2016) state if RMSEA is ≤ 0.05 then, it indicates good fit and if RMSEA is ≤ 0.08 it indicates good fit with reasonable errors. Other recommendations generally advise that if $RMSEA \leq 0.08$ then a model has good fit and, if $0.08 < RMSEA \leq 0.1$

then it indicates adequate fit (Carlback & Wong, 2018; Shadfar & Malekmohammadi, 2013). The confirmatory analysis revealed that RMSEA value is .067 so SAS has good fit of model.

Since RMR has bias, SRMR used instead of RMR. SRMR value of the confirmatory factor analysis result is 0.0746. SRMR value below ≤ 0.08 indicates good fit (Carlback & Wong, 2018; Kline, 2011 as cited in; Kaya & Altinkurt, 2018; Vassallo & Saba, 2015).

Kline (2011) mentions that most used fit indices (GFI, AGFI, NFI, NNFI, CFI and IFI) should be ≥ 0.85 (cited in; Kaya & Altinkurt, 2018; Vassallo & Saba, 2015). Other researchers indicate AGFI, GFI and CFI values should be ≥ 0.80 (Byrne & Campbell, 1999 as cited in Nayir, 2013). However, GFI is affected by sample size and for that reason, AGFI is developed. $GFI \geq 0.85$ and $AGFI \geq 0.80$ is accepted as good fit (Sica & Ghisi, 2007). On the other hand, AGFI also is sensitive to sample size. For that reason, it is advised by researchers to disregard them. Yet papers still indicate both values. Reason for that is not for their importance but historical values. $TLI \geq 0.85$ indicates good fit and > 0.8 mediocre fit (Carlback & Wong, 2018; Shadfar & Malekmohammadi, 2013). The confirmatory analysis revealed that IFI value is 0.806; GFI value is .852; CFI value is .800; NNFI (TLI) value is 0.773; AGFI value is .815 so SAS has good fit of model.

4. DISCUSSION and CONCLUSION

SAS item discrimination values show that it might be used to measure the attitude toward science of college students since all the items yielded significant results between up-down%27. On the other hand, a detailed analysis of t-test values might reveal the facts lying beneath. For example, the highest t value (26,213) of item9 implies that students who have high positive attitudes enjoy the experiments, however; the value of items11 and 21 (7,381 and 6,093) imply students wait for confirmation on their experiment results or expect more guidance during experiments. Fin (2012) states perceived learning increases as students get feedback from instructors, thus; the expectation of students is meaningful and expected in this context. Similar reports indicate that instructor confirmation has a positive effect on cognitive learning (Schrodt, Witt, Turman, Myers, Barton & Jernberg, 2009). Therefore, instructors need to interact with students and help them in cognitive development.

Students have a tendency of thinking that feelings might assist a scientist. Low t value (4,944) of item 7 already implies that both upper-lower (%27) students think that facts may have subjective aspects. Perhaps, students think that scientific facts may change, and some unscientific factors act as catalysts for that change. In fact, t value (5,767) of SAS1 already hints that both upper-lower (%27) students have a tendency of the idea that facts might be subjective. From a different aspect, that approach might be seen through t value (6,588) of SAS6. That item indicates faith plays an important role in students' ideas about science's role. Conflicts between the dimensions may create gaps for the students on the nature of science. As a result, students struggle between scientific facts and faith. Since faith requires believing in what is being told without question, it may also create a barrier toward scientific approach and hence, students may close themselves to new ideas. Students who have high positive attitudes toward science are also open to new ideas and t value (22,715) of SAS12 already reveals this. One of the purposes of science courses is learning how to distinguish faith from science and instructors should make students aware that; two concepts actually are not related. In fact, faith and science have different roles and do not need to cut out each other's way. For instance, a high value of SAS19 ($t=20,686$) clarifies that students who have higher positive attitudes also acknowledge that scientific knowledge is essential as it is related to life itself. Embracing this idea might eventually increase positive attitudes toward science. Similarly, Nuhoglu (2008) who developed a science attitude scale in Turkey mentions that one of the factors is "new knowledge and using it" and Ajzen (2005a) points out that changing the attitudes and behaviors may be achieved through changing beliefs.

For a cross-cultural analysis, Scientific Attitude Inventory II (SAI-II) was also analyzed since some statements are similar to the statements in SAI-II. However, researchers did not provide an item t%27up-down values thus a comparison could not be made (Moore & Foy, 1997). Similarly, another well-known scientific attitude scale TOSRA, which has seven dimensions, also did not provide such an approach. On the other hand, the author states that having a low score on one of the dimensions should not concern instructors because this information might be valuable in facilitating to identify profiles and creating solutions. One of the most important aims of science education is considered to develop a positive attitude towards science (Fraser, 1981).

Motivation towards science has a long-term effect on science learning and it is affected by different things such as curriculum structures. For that reason, Foley & McPhee (2008) investigated the effect of different curriculum approaches on their study and argued that students' experiences might be affected by different curriculum structures. Kurnaz and Yigit (2010) report that Turkish students have tended to develop a negative attitude towards science since 2005. Thus, it is important to assess the changes of attitudes caused by curriculum changes in Turkey (TTK, 2017). Although this research was done on local scale, it might be said that science attitude affects the scientific approach and the scientific approach is the same all around the world. Moreover, attitudes might be affected by a similar insight context whose effect may yield similar results. Perhaps through such studies, identifying and creating solutions will be possible. Researchers and teachers might use the developed scale and observe their students', attitudes, use interventions and may offer solutions. Then, perhaps understanding students' attitudes on different dimensions may also offer solutions for long term aims.

An advantage of this study is seen as using parallel analysis to confirm eigenvalues in identifying dimensions. Since the traditional factor analysis determination procedure is based on eigenvalue of 1 then, the obtained number of factors may not be accurate. It is believed through such analysis such cases are avoided and true factor structures are determined in creating the SAS (Hayton, Allen & Scarpello, 2004). Developing science attitude scales is important, even restudying the existing ones may provide useful information. For that reason, researchers either create their own scales or develop the existing ones (Moore & Foy, 1997). Having an excellent internal consistency ($\alpha=.953$) and a consisted structure (one factor structure) measuring the related domain, it is believed that SAS is a valuable contribution to the science education field.

Validity must be considered each time when an instrument is used since the instrument was validated for a sample or population but was not validated for another sample or population i.e. structure may show varying results from a sample to another. Validity is not property of a scale but it means as an instrument of interpretation. There are arguments on cut off values of fit indices since they may lead the decision of an acceptable model to be rejected. Thus, it is important for researchers to conduct the analyses and use their own judgments with respect to obtained values. Values of fit indices help to understand the structure of a model and thinking all the fit values together will provide a better decision. In other words, fit indices will help to understand relationships of the items among each other and within the model structure. Fit indices should confirm the model but it should not be used for championing the model in every possible indices which will cause an artificially approved model. It is advised that once items and factors make sense in the theoretical aspect of the researcher then, the decision will be based on that (Knekta, Runyon & Eddy, 2019). Hu & Bentler (1995) argue decisions based on fit indices and reminds that strictly depending on fit indices values may result in rejecting true models, especially for small sample sizes such as 250 or 500. GFI and AGFI tend to increase when sample size increase same thing could be also said for RMR and RHO (Anderson & Gerbing, 1984). NFI results could be problematic if sample size is < 200 thus usage of NNFI (TLI) is recommended. However, it is also noted that even NNFI could still indicate poor fit if

sample size is not enough. Similar arguments are proposed for TLI since it could report poor fit due to similar reasons. For that reason, suggestions for TLI could be as low as 0.80 and, for parsimony fit indices (PNFI) values ≥ 0.5 indicates good fitness (Hooper, Coughlan & Mullen, 2008). The confirmatory analysis revealed that PNFI value is .583 so SAS has good fit of model. Although there are other authors who conclude that with sample size $N=200$ a reasonable estimate could be obtained for CFI and TLI, still researchers are warned since decisions strictly based on CFI could also cause wrong decisions because it also depends on sample size and hence rejection of fit model. For example, a correct model simulated with a sample $N=200$ and CFI value turned out to be .611 (poorly fitting model). It is noted that even with a relatively large sample size ($N=500$) a conventional cut off value of TLI may cause a correct model to be rejected (Shi, Lee, Maydeu-Olivares, 2018). Hu & Bentler (1999) notes sample size ≤ 250 could cause problems in Maximum Likelihood (ML) analysis. Thus, warns researchers to be cautionary on evaluation on model fit evaluation. Questions (items) assessing the same target or different items having nearly same meaning with different words may be the cause of correlated errors which in fact, may cause the correlate error terms (Meyer, n.d.). Since SAS has correlated error terms, this also concludes the idea that model could provide a multidimensional aspect if provided with enough number of items targeting the domain of interest. Be that as it may, Ellis (2017) states that if p value is < 0.05 and $0.05 < RMSEA < 0.08$ then, null hypothesis is not exactly true but model has acceptable fit. Although generally accepted indice values are ≥ 0.80 for fit indices, a proposed common guideline for indice values follows as; very good fit ≥ 0.90 ; adequate but marginal fit $\geq 0.80-0.89$; poor fit $\geq 0.60-0.79$; very poor fit > 0.60 (Planing, 2014).

As a final thought, it is believed that the developed SAS might be used in different regions/states to compare the results and validate its purpose. For that reason, SAS with different samples is welcome. With this aim, both the created SAS in the original language and an English version of SAS are given in the appendixes.

ORCID

Adem AKKUŞ  <https://orcid.org/0000-0001-9570-3582>

5. REFERENCES

- Adesoji, F. A. (2008). Managing students' attitude towards science through problem-solving instructional strategy. *The Anthropologist*, 10 (1), 21-24.
- Ajzen, I. (2005a). Behavioral interventions based on the theory of planned behavior: Brief description of the theory of planned behavior. Retrieved from <http://people.umass.edu/ajzen/pdf/tpb.intervention.pdf> (accessed on 5 February 2019)
- Ajzen, I. (2005b). Constructing a theory of planned behavior questionnaire: Brief description of the theory of planned behavior. Retrieved from <http://people.umass.edu/ajzen/pdf/tpb.measurement.pdf> (accessed on 5 February 2019)
- Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155–173. DOI: 10.1007/bf02294170
- Anthoine, E., Moret, L., Regnault, A., Sébille, V., & Hardouin, J.-B. (2014). Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes*, 12, 2. DOI:10.1186/s12955-014-0176-2
- Bernardez, R.Q. (1982). *Factors affecting attitudes to laboratory work*. Unpublished Master Thesis, Saint Louis University, Baguio, Philippines. [Abstract]
- Brinkman, W-P. (2009). *Design of a questionnaire instrument, handbook of mobile technology research methods*. ISBN 978-1-60692-767-0, pp. 31-57 Netherlands: Nova Publisher
- Browne, M.W. & Cudeck, R., 1993. *Alternative ways of assessing model fit*. In: Bollen, K.A. and Long, J. S. (Eds.) *Testing structural equation models*, Beverly Hills, CA: Sage

- Byrne, B. M. & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, 30, 557 - 576. DOI: <https://doi.org/10.1177/0022022199030005001>
- Cabrera-Nguyen, P. (2010). Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research. *Journal of the Society for Social Work and Research*, 1(2), 99-103.
- Can, M., & Şahin, Ç. (2015). Okul öncesi öğretmen adaylarının fene ve fen öğretimine yönelik tutumlarının incelenmesi [Investigating Prospective Kindergarten Teachers' Science and Science Teaching Attitudes]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 15 (2), 13-26. DOI: [10.17240/aibuefd.2015.15.2-5000161311](https://doi.org/10.17240/aibuefd.2015.15.2-5000161311)
- Carlbäck, J. & Wong, A. (2018). *A study on factors influencing acceptance of using mobile electronic identification applications in Sweden*. Retrieved from <http://www.diva-portal.org/smash/get/diva2:1214313/FULLTEXT01.pdf> (accessed on 03 April 2019)
- Cheung, D. (2007). Confirmatory factor analysis of the attitude toward chemistry lessons scale. Paper presented at 2nd NICE symposium, Taipei, Taiwan, July 30-31, 2007.
- Chinda, T., Techapreechawong, S., & Teeraprasert, S. (2012). *An investigation of relationships between employees' safety and productivity*. Retrieved from http://www.ppml.url.tw/EPPM/conferences/2012/download/SESSON4_A/10%20E145.pdf (accessed on 12 October, 2019)
- Coll, R.L., Dalgety, J. & Salter, D. (2002). The development of the chemistry attitudes and experiences questionnaire (CAEQ). *Chemistry Education Research and Practice in Europe*, 3(1), 19-32.
- Demirbaş, M. (2009). The relationships between the scientist perception and scientific attitudes of science teacher candidates in Turkey: A case study. *Scientific Research and Essays*, 4(6), 565-576.
- Deshpande, L. (2004). Challenges in measurement of scientific attitude. Paper presented at *epiSTEME-1: An International Conference to Review Research on Science Technology and Mathematics Education* (137-138), Goa, India, December 13-17, 2004.
- Dönmez, F., & Azizoğlu, N. (2010). Investigation of the students science process skill levels in vocational schools: a case of Balıkesir. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, 4 (2), 79-109.
- Ellis, J.E. (2017). Factor analysis and item analysis. *Applying Statistics in Behavioural Research* (pp. 11-59). Retrieved from https://www.applyingstatisticsinbehaviouralresearch.com/documenten/factor_analysis_and_item_analysis_version_11_.pdf (accessed on 22 December 2018)
- Evrekli, E., İnel, D., Balım, A. G., & Kesercioğlu, T. (2009). The attitude scale of constructivist approach for prospective science teachers: a study of validity and reliability. *Journal of Turkish Science Education*, 6(2), 134-148.
- Field, A. (2013). *Discovering statistics using ibm spss statistics* (4th Edition). London: SAGE
- Finn, A.N. (2012) Teacher use of prosocial and antisocial power bases and students' perceived instructor understanding and misunderstanding in the college classroom. *Communication Education*, 61(1), 67-79, DOI: [10.1080/03634523.2011.636450](https://doi.org/10.1080/03634523.2011.636450)
- Foley, B., & McPhee, C. (2008). Students' attitudes towards science in classes using hands-on or textbook based curriculum. *AERA*, 1-12.
- Fraser, B.J. (1981). *Test of science-related attitudes (TOSRA) handbook*. Victoria: Allanby
- Garson, G.D., 2006. *Structural equation modelling*. North Carolina: G. David Garson and Statistical Associates Publishing
- Hayton, J.C., Allen, D.G & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191-205. DOI: [10.1177/1094428104263675](https://doi.org/10.1177/1094428104263675)

- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 2 (1), 104-121. DOI:10.1177/109442819800100106
- Hinkin, T. R., Tracey, J. B., & Enz, C. A. (1997). Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research*, 21(1), 100-120. DOI:10.1177/109634809702100108
- Hof, M.W. (2012). *Questionnaire Evaluation with Factor Analysis and Cronbach's Alpha: An Example*. Retrieved from <http://www.let.rug.nl/nerbonne/teach/rema-stats-meth-seminar/student-papers/MHof-QuestionnaireEvaluation-2012-Cronbach-FactAnalysis.pdf> (accessed on 02 May 2016).
- Hooper, D., Coughlan, J. & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, 6 (1), 53 – 60. DOI: 10.21427/D7CF7R
- Hu, L.-T., & Bentler, P. M. (1995). *Evaluating model fit: in Structural Equation Modeling* Ed. Rick H. Hoyle. London: Sage Publications
- Hu, L.-T., & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. DOI: <https://dx.doi.org/10.1080/10705519909540118>
- Johanson, G.A., & Brooks, G.P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement*, 70(3), 394-400. DOI: 10.1177/0013164409355692
- Johnson, R.L. & Morgan, G.B. (2016). *Survey scales: Investigating scale quality*. New York, NY: The Guilford Press.
- Kalaycı, Ş. (2010). Spss uygulamalı çok değişkenli istatistik teknikleri. (5. Baskı) [SPSS applied various statistical techniques (5th Edition)]. Ankara: Asil Yayın Dağıtım Ltd. Şti.
- Kaya, Ç. & Altinkurt, Y. (2018). Öğretmenlerin psikolojik sermayeleri ile tükenmişlik düzeyleri arasındaki ilişkide psikolojik ve yapısal güçlendirmenin rolü [Role of Psychological and Structural Empowerment in the Relationship between Teachers' Psychological Capital and Their Levels of Burnout]. *Eğitim ve Bilim*, 43 (193), 63-78, DOI: <http://dx.doi.org/10.15390/EB.2018.6961>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd edition) New York: The Guilford Press
- Knekta, E., Runyon, C., & Eddy, S. (2019). One Size Doesn't Fit All: Using Factor Analysis to Gather Validity Evidence When Using Surveys in Your Research. *CBE—Life Sciences Education*, 18 (1), 1-17. DOI: <https://doi.org/10.1187/cbe.18-04-0064>
- Kurnaz, M.A. & Yigit, N. (2010). Physics attitude scale: Development, validity and reliability. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, 4 (1), 29-49.
- Korkmaz, Ö., Şahin, A. & Yeşil, R. (2011). Bilimsel araştırmaya yönelik tutum ölçeği geçerlik ve güvenilirlik çalışması [Study of Validity and Reliability of Scale of Attitude towards Scientific Research]. *Elementary Education Online*, 10 (3), 961-973.
- Lovelace, M. & Brickman, P. (2013). Best practices for measuring students' attitudes toward learning science. *CBE-Life Sciences Education*, 12(4), 606-617. DOI: 10.1187/cbe.12-11-0197
- Meyer, J. (n.d.). Correlated errors in confirmatory factor analysis. Retrieved from <https://www.theanalysisfactor.com/correlated-errors-in-confirmatory-factor-analysis> (accessed on 17 September 2019).
- Moore, R.W. & Foy, R.L.H. (1997). The scientific attitude inventory: A revision (SAI II). *Journal of Research in Science Teaching*, 34(4), 327-336.
- Nayir, F. (2013). “Algılanan örgütsel destek ölçeğinin” kısa form geçerlik güvenilirlik çalışması [“Perceived Organizational Support Scale”- Short Form Validity-Reliability Study]. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 28, 89-106.

- Nuhoglu, H. (2008). The development of an attitude scale for science and technology course. *Elementary Education Online*, 7(3), 627-639.
- Pedroso, R., Zanetello, L., Guimaraes, L., Pettenon, M., Goncalves, V., Scherer, J., Kessler, F., & Pechansky, F. (2016). Confirmatory factor analysis (CFA) of the crack use relapse scale (CURS). *Archives of Clinical Psychiatry*, 43 (3), 37-40.
- Planing, P. (2014). Innovation Acceptance: The Case of Advanced Driver-Assistance Systems: *Quantitative research approach* (pp. 230-231). Stuttgart: Springer
- Sica, C. & Ghisi, M. (2007). The Italian versions of the Beck Anxiety Inventory and the Beck Depression Inventory-II: Psychometric properties and discriminant power. In M.A. Lange (Ed.), *Leading - Edge Psychological Tests and Testing Research* (pp. 27-50). New York: Nova
- Schrodt, P., Witt, P.L., Turman, P.D., Myers, S.A., Barton, M.H & Jernberg, K.A. (2009) Instructor credibility as a mediator of instructors' prosocial communication behaviors and students' learning outcomes. *Communication Education*, 58(3), 350-371, DOI: [10.1080/03634520902926851](https://doi.org/10.1080/03634520902926851)
- Serin, O., & Mohammadzadeh, B. (2008). The relationship between primary school students' attitudes towards science and their science achievement (sampling: Izmir). *Journal of Educational Sciences*, 2 (6), 68-75.
- Shadfar, M. & Malekmohammadi, I. (2013). Application of Structural Equation Modeling (SEM) in restructuring state intervention strategies toward paddy production development. *International Journal of Academic Research in Business and Social Sciences*, 3 (12), 576-618. DOI: [10.6007/IJARBS/v3-i12/472](https://doi.org/10.6007/IJARBS/v3-i12/472)
- Shah, Z.A. & Mahmood, N. (2011). Developing a Scale to Measure Attitude towards Science Learning among School Students. *Bulletin of Education and Research*, 33 (1), 71-81.
- Shi, D., Lee, T., & Maydeu-Olivares, A. (2018). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, 79(2), 310-334. DOI: <https://doi.org/10.1177%2F0013164418783530>
- Tortop, H. S. (2013). Bilimsel Alan Gezisi Tutum Ölçeği Adaptasyon Çalışması [Adaptation Study of Attitude Scale towards Scientific Field Trips]. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, 2(1), 228.
- TTK. (2017, July 18). *Müfredatta yenileme ve değişiklik çalışmalarımız üzerine [On curriculum revision and changes work]*. Retrieved from https://tkb.meb.gov.tr/meb_iys_dosyalar/2017_07/18160003_basin_aciklamasi-program.pdf (accessed on 25 August 2018).
- Watkins, M. W. (2000). *Monte carlo PCA for parallel analysis [computer software]*. State College, PA: Ed & Psych Associates.
- Wong., M. & Lian, S. (2003). Development of a self-efficacy scale for assessing secondary school students' science self-efficacy beliefs. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.472.8479> (accessed on 19 October 2014)
- Vassallo, M., & Saba, A. (2015). Does money for grocery expenditure sway Italian consumers' motivational values in predicting Attitude towards eco-sustainable food products? *Contemporary Management Research*, 11(1), 3-22. DOI: [doi:10.7903/cmr.13840](https://doi.org/10.7903/cmr.13840)

6. APPENDIX

Table A1. Bilimsel Tutum Ölçeği

Madde	Bilimsel Tutum	Kesinlikle Katılmıyorum	Katılmıyorum	Kararsızım	Katılıyorum	Kesinlikle Katılıyorum
1	Bilimsel gerçekler değişmez	1	2	3	4	5
2	Bilim adamları gerçeğin/fikirlerin değişebileceğine inanırlar/kabul ederler	1	2	3	4	5
3	Bilim adamları sorularının cevaplarını her zaman bulamazlar	1	2	3	4	5
4	Bilim adamları birbirlerinin çalışmalarını eleştirmemelidirler	1	2	3	4	5
5	Bilimsel çalışmalar bilim adamları içindir	1	2	3	4	5
6	Dinle çatışan konular çalışılmamalı/önemsenmemeli	1	2	3	4	5
7	Bir bilim adamının sahip olduğu en önemli araç hisleridir	1	2	3	4	5
8	Bilimsel gelişmeler daha sağlıklı yaşam sürmemizi sağlar	1	2	3	4	5
9	Deney yapmak derslerden daha zevklidir	1	2	3	4	5
10	Bilimsel keşifler faydadan çok zarar veriyor	1	2	3	4	5
11	Hocanın anlatmasındansa deney yaparak gerçekleri bulmayı tercih ederim	1	2	3	4	5
12	Farklı fikirleri hoş karşılarım	1	2	3	4	5
13	Fen dersleri zaman kaybıdır	1	2	3	4	5
14	Fen konuları zevksizdir	1	2	3	4	5
15	Deney yapmaktansa teorik bilgiler daha faydalıdır	1	2	3	4	5
16	Fen deneylerine daha az vakit verilmeli	1	2	3	4	5
17	Deneyler grup çalışmasıyla daha zevkli geçer/geçiyor	1	2	3	4	5
18	Bilimin temel amaçlarından biri yeni ilaçlar ve tedaviler bulmaktır	1	2	3	4	5
19	Yaşamı etkilediğinden İnsanlar bilimsel gerçekleri anlamalı	1	2	3	4	5
20	Bilim bir şeyin nasıl olduğunu açıklamaya çalışmaktır	1	2	3	4	5
21	Bilimsel çalışma bana zor gelir	1	2	3	4	5

Table A2. Scientific Attitude Scale

Item	Scientific Attitude	Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
1	Scientific facts do not change	1	2	3	4	5
2	Scientists acknowledge/accept that facts may change	1	2	3	4	5
3	Scientists cannot always find the answers	1	2	3	4	5
4	Scientists should not criticize each other's work	1	2	3	4	5
5	Scientific works are for scientists	1	2	3	4	5
6	Topics contradicting with religion should not be studied/cared	1	2	3	4	5
7	The most important tool for a scientist is her/his feelings	1	2	3	4	5
8	Scientific progress helps us to have more healthy life	1	2	3	4	5
9	Doing experiments is more fun than having lectures	1	2	3	4	5
10	Scientific progress outputs harm more than good	1	2	3	4	5
11	I prefer to find facts rather than told by the instructor	1	2	3	4	5
12	I welcome different ideas	1	2	3	4	5
13	Science courses are waste of time	1	2	3	4	5
14	Science courses are tasteless	1	2	3	4	5
15	Theoretical knowledge is more helpful than experimenting	1	2	3	4	5
16	Science course hours must be reduced	1	2	3	4	5
17	Experiments are more fun with group works	1	2	3	4	5
18	One of the main aims of science is to find new cures	1	2	3	4	5
19	People should understand scientific facts since it affects life	1	2	3	4	5
20	Science is trying to explain things	1	2	3	4	5
21	Scientific works are baffling for me	1	2	3	4	5