A Review of Attitude Scales Developed in Turkey Between 2002-2018 Regarding the Scale Development Process

Gül Güler [iD] [1]
Istanbul Aydın University

Cansu Ayan [iD] [2]
Ankara University

Abstract

The purpose of this study is to determine the degree to which the process steps to be taken into account in developing attitude scale for use in education and psychology are met. In this way, the problems encountered in the literature about the scale development process will be identified. In addition, this study is considered important in terms of being guiding and informative for the next attitude scale development studies. The journals indexed in ULAKBIM and Turkish educational journals that can be accessed full texts electronically are included in the scope of the study. In this context, 112 attitude scales development studies conducted in the field of education between 2002 and 2018 in Turkey were examined. This study is a qualitative study as it is conducted by taking into consideration the studies developing attitude scale as well as the points that ought to be considered while developing attitude scale. Articles were examined according to pre-determined criteria, and the frequency values were obtained for each criterion. Attitude scale development studies within the scope of this study have attempted to determine the points which are often inaccurate or incomplete in the literature. It is recommended that the points highlighted in the findings of this study are taken into consideration and the scale development process should be considered as an important and rigorous process.

Keywords: Attitude, attitude scale, scale development procedure.

[1]*Corresponding Author*: Assist. Prof. Dr., Education Faculty, Basic Education Department, E-mail: gulyuce@aydin.edu.tr, https://orcid.org/0000-0001-8626-4901
[2]Dr., Faculty of Educational Sciences, Departmant of Educational Sciences, E-mail: cayan@ankara.edu.tr, https://orcid.org/0000-0002-0773-5486

The concepts that are dealt with in the fields of education and psychology are mainly related to affective characteristics. Unlike concrete characteristics, affective characteristics due to their nature cannot be directly observed. Therefore, in studies investigating these characteristics, we use indirect measurement methods when they need to be measured (Anastasi, 1988; Baykul, 2000; Özgüven, 2011). In indirect measurements, a variable is measured by other variables. For this purpose, individuals are compared by means of a number of stimuli that will disclose the nature of the psychological characteristics concerned (Kilmen, 2017; Turgut and Baykul, 2010). The most common way to use this in education and psychology is through psychological measurement tools. Psychological measurement tools generally consist of items that exemplify a group of behaviours that are considered to be indicative of the psychological characteristics that are to be measured. Therefore, it is accepted that the psychological trait in a question is measured with the help of the measurement instrument, based on the responses given by the individual to these indicators (Cronbach,1990; Özgüven, 2011).

The accuracy, generalizability, and functionality of the findings obtained from psychological measurement tools are directly proportional to the reliability and validity of these tools (Erkuş, 2007). The accuracy and reliability of the results as obtained by using the measurement results whose validity and reliability are suspected, and which do not follow the measurement development stages meticulously and correctly are also debatable (Crocker and Algina, 1986). In the social sciences, it is often observed that inconsistent results are obtained in different studies where the same variables are measured. It is discussed that one of the possible causes of this situation may be the use of different measuring tools (Hinkin, 1995; Kaya-Uyanık, Güler, Taşdelen-Teker and Demir, 2017; Schriesheim, Powers, Scandura, Gardiner and Lankau, 1993).

One of the most important points of developing a valid and reliable psychological measurement instrument is the fact that the test developer has a good knowledge of the psychological structure. Researchers who lack sufficient knowledge about the definition, characteristics, sub-dimensions, and possible indicators of the psychological structure to be measured will also have difficulty developing the scale in order to measure the construct, and moreover will be more prone to making mistakes when determining the indicators (Schultz and Schultz, trans. 2007). Erkuş (2012) refers to this fact, noting that it is not appropriate to develop a scale either without knowing the concept to be measured, or without knowing the measurement process despite knowing the concept to be measured. Therefore, it is possible to consider the scale development process as a difficult and specialized job that requires mastery of both the psychological structure concerned and the field of measurement and evaluation.

One of the most frequent affective characteristics to be measured in both education and psychology research is attitude. Attitude can be expressed as the individual's tendency to orient his/her behaviour, thoughts and emotions related to the

psychological object, and to be against a particular thing or an individual (Turgut, 1977). According to Tavşancıl (2005), attitude is an emotional and mental preparation that is formed as a result of life experiences, and which has the power to influence or direct the behaviours of the individual in relation to all the objects and situations. One of the important developments in measuring attitude is the study of Thurstone. L. L. Thurstone (1929) introduced a scaling approach to the measurement of attitudes. With this method, he developed a scale by using the judgments of experts and reflecting the positive and negative emotions related to the situation. Likert (1932) used a five-fold rating scale from positive to negative, which is a slightly different approach. He also introduced the first examples of Likert type scales, which still are frequently used.

Regardless of which psychological concept is of concern, the scale development process has a number of stages to follow. Although many of these stages have been identified by a number of scientists (Coaley, 2010; Cohen and Swerdlik, 2010; Crocker and Algina, 1986; Erkuş, 2012; Murphy and Davidshofer, 2005; Rust and Golombok, 1997; Tezbaşaran, 2008; Turgut, 1977) it appears that the basic stages and the procedures to be carried out are essentially the same. Tezbaşaran (2008) discussed the attitude scale development stages under three main headings: regulating the trial form of the scale, carrying out the trial application, and analysing the data obtained from the trial application. He listed the works to be carried out under the headlines below:

- Determination of the scope of attitude,
- Identification of appropriate observable indicators in conformity with the scope,
- Preparation of scale items,
- Preparation of directives,
- Determination of the order of items in scale,
- Conducting pre-examination,
- Implementation of the trial application,
- Scoring the answers given to items,
- Calculation of individuals' raw scores,
- Determination of features of the raw score distribution,
- Determination of the characteristics of item scores distribution,
- To evaluate the items and scale (item analysis, validity, reliability analysis, factor analysis etc.),
- To finalize the scale.

Having examined studies dealing with scale development in the literature, it was observed that both the number of studies had increased annually, and that that there were significant technical problems in the existing studies. Misapplications in the literature provide a bad example, whereby some problems become chronic, causing other studies to be repeated in the same way. Determining these problems and deficiencies is important in terminology of encouraging future studies. Moreover, it is possible to come across studies examining the stages to be followed in scale

development or adaptation in literature. These studies examine scale development and adaptation studies together (Acar-Güvendir and Özer-Özkan, 2015; Çüm and Koç, 2013; Erkuş, 2007) or separately (Boztunç-Öztürk, Eroğlu and Kelecioğlu, 2014). Furthermore, no study focusing on scales developed by a specific psychometric structure were found during the examination of the scale development/adaptation stages in general. Attitude scales are one of the most common psychological measurement tools in the literature. In this context, the psychological structure of the attitude and the correct determination of the indicators that represent it are considered to be important. The fact that a wide range of attitude scales exist that are virtual of little use is a sign of a serious loss of labour and time. The fact that no study thoroughly examining the attitude scale studies developed in the fields of education and psychology in the literature presently exists, and the fact that revealing the problems in this field will be an important step for the elimination of the problems constitutes the necessity of this study.

Purpose

The purpose of this study is to determine the degree to which the process stages to be taken into account in developing an attitude scale for use in education are met. In line with this purpose, we attempted to answer the following questions based on the articles reviewed:

1. Introductory Information
    - What are the attitude issues that are dealt with to develop the scale?
    - What is the number of items being tested, the number of items in the last version of the scale and the size of the sample group when developing the scale for each study?
    - What are the category numbers in graded expressions?

2. Theoretical Section
    - Has the psychological structure to be measured been defined in detail?
    - Has the operational definition of psychological construct been made?
    - Has the operational definition been made correctly?

3. Item Writing and Trial Application Section
    - Have items been written in accordance with the principles of item writing?
    - Is the distribution of positive/negative items balanced?
    - Is the distribution of the items related to cognitive, affective and dynamic components of the attitude balanced?
    - Are the rating expressions used with the written items suitable for each other?
    - Has an expert opinion been given (i.e. by a measurement and evaluation expert, a Turkish linguist, and subject matter expert)?

4. Reliability Section
    - Is there a study conducted on the reliability of the final scale?

- Which reliability determination methods have been used to prove the reliability of the scales?
- Are reliability studies appropriate?

5. Validity Section
   - Have the validity studies of the scales been carried out?
   - What validity methods have been used to prove the validity of the scales?
   - If the study conducted is a factor analysis, have the KMO and Barlett test results, which are a prerequisite for factor analysis, been included?
   - Have the results of the factor analysis been reported in an appropriate manner?
   - If a criterion validity study was conducted, is the criteria used appropriate?
   - Has the information given about the psychometric properties of the criterion been used?
   - Are validity studies of scales suitable?
   - Have the test statistics related to the distribution of the scales been included?

Method

In this section, information about the research model, sample, data collection instruments and data analysis process has been reported. The ethical committee approval is not compulsory for this research because it was sent to our journal before 01.01.2020.

Research Model

This study is a document investigation study being in scope of qualitative researches as it is conducted by taking into consideration the studies developing attitude scale as well as the points that ought to be considered while developing attitude scale. Articles were examined according to pre-determined criteria, and the frequency values were obtained for each criterion.

Universe-Sample

This study aims to examine the attitude scale development studies conducted in the field of education between 2002 and 2018 in Turkey. In order to provide national information and document access services as a first step, journals were scanned at Turkish Academic Network and Information Center (ULAKBIM), an institute founded by The Scientific and Technological Research Council of Turkey (TÜBİTAK). Educational journals, whereby full texts can be reached via electronic media, were also scanned in order to provide national information and document access services as the first step. A total of one hundred and twelve articles were found.

Data Collection Instruments

The data of the study were collected using a coding list which is developed in Tavşancıl, Güler and Ayan's study (2014). The coding list is based both on the points to be considered while developing the attitude scale and the purpose of the research,

as well as on the relevant literature. It is then submitted to the opinion of three measurement and evaluation specialists who were academicians in the field of measurement and evaluation. Having examined the developed coding list, it has been seen that the list consists of three basic parts. The initial section deals with the preliminary information about the study is questioned for main items questioning the stages to be followed with regards to the four main headings. The second section contains answers that are graded in the forms of *yes*, *no*, and *no information*. The third section features questions related to the validity and reliability evidence used in the studies. Having examined the second section of the coding list in detail, four main headings come to the foreground: theoretical and operational definitions, item writing and trial application, reliability, and validity. Each section contains the relevant items that question the points that should be done in that section.

A separate coding list was used in answering the question "have the items been written in accordance with the principles of item writing?" (one of the items in the coding list), whereupon each item was looked at to see if it matched with the criteria found within this list. The criteria of which the non-conforming items were in violation were stated, and direct quotations were formed from the sample items.

Data Analysis

The data were subjected to content analysis, a type of analysis used in qualitative research. The categorical analysis method, which is one form of content analysis, was applied, and the frequencies of each category were calculated.

In the categorical analysis, there are two ways to follow the category system. The first one is the Theoretical Category Formation Process, whilst the second is the Practical Category Formation Process (Tavşancıl and Aslan, 2001). In clearer terminology, the categories in the coding process can be certain in the beginning because it starts from a theoretical basis. They can also be created in the process by the researcher as the materials to be examined begin to be examined. This, in other words, is called category formation with either a deductive or an inductive approach. Sometimes these two processes can be used together. Although the researcher begins to codify with a ready-made category system on a theoretical basis, they can change the coding system as the materials are examined (Bilgin, 2006; Tavşancıl and Aslan, 2001). The researchers formed the categories formed within the scope of this study by making additions and subtractions in the coding process. This began from a theoretical basis that clarifies the points to be considered in developing an attitude scale. Thus, both deduction and induction methods were applied. Coding lists were used for coding operations. Each researcher read all the items included in the research and made the appropriate coding for each of the items. Finally, the frequency values for each point were calculated and reported.

The reliability of the content analysis is particularly dependent on the coding process. If the process of category determination is meticulously carried out, the possibility of working with high reliability is quite high. The fact that the interpretations of the categories do not change from researcher to researcher, or at two

different times, provides reliability as a condition of objectivity (Tavşancıl and Aslan, 2001). The inter-researcher reliability was calculated as a proof of reliability in this study, and the consistency of the codes made by two different researchers was then examined. For the calculation of intra-researcher reliability, the following formula [generally used by Miles and Huberman (1994)] was used to determine the reliability of content analysis studies. The percentage of fit between the rates is expected to be higher than 70% (Tavşancıl and Aslan, 2001). Reliability = number of compromise / (number of compromises + number of non-compromises). In this context, the intercoder reliability was found to be 0.87.

Results

In this part, results which were obtained from document analysis were presented.

Distribution of the Attitude Topics Dealt within the Articles

The attitude topics as discussed in the attitude scale development studies were examined, and their corresponding frequencies were calculated and listed. The results are presented in Table 1.

Table 1

Distribution of the Attitude Topics Dealt Within the Articles

	Attitude Topics	**f**	**%**
	Science and Technology	10	
	Planning and Evaluation for Teachers	1	
	Information Networks and Communication	1	
	Painting	2	
	Lab Courses	2	
	School Experience	2	
	Piano	1	
	English	2	
Course-Oriented	Mathematics	4	
	Biology	3	
	Music	4	
	Geography	1	
	Chemistry	2	
	Turkish	3	
	Geometry	1	
	Media Literacy	1	
	History of Turkish Language Education	1	
	Total	41	36.61
	Teaching Profession	2	
	Career choice	1	
Occupation-Oriented	Music Pedagogy	1	
	Biology Pedagogy	1	
	Total	5	4.46

(continued)

Table 1 (continue)

	Attitude Topics	f	%
Technology-Oriented	Internet	1	
	Distance Education	1	
	Mobile Learning	1	
	Digital Technology	1	
	Computer	3	
	Use of technology during the course	1	
	Information technology	2	
	Information and communication technology	1	
	Auxiliary Technology	1	
	Total	12	10.71
A Teaching Method-Technique-Oriented	Concept Mapping	1	
	Mind Mapping	1	
	Constructivist Approach	1	
	Problem-Based Teaching	1	
	Modular Teaching	1	
	Student-Centred Teaching Methods and Techniques	2	
	Proof and Proving in Math	1	
	Reading Scientific Texts	1	
	Using Models in Science & Technology Courses	1	
	Total	10	8.93
Other Education-Related Elements-Oriented	Cheating Behaviour	1	
	Science Experiments	1	
	Homework	1	
	In-Class Use of Equipment	1	
	Reading Habits	2	
	Inspector	1	
	Undesirable In-Class Behaviours	1	
	Integration of Science and Art Issues	1	
	Using Graphics	1	
	Educational Games	1	
	Absence	1	
	Dictionary	1	
	Augmented Reality Application	1	
	Using English on the Internet	1	
	Family Involvement	1	
	Rating Key	1	
	Writing-Oriented	1	
	Grammar-Oriented	1	
	Turkish Language Activities	1	
	Listening-Oriented	1	
	Educational Research-Oriented	1	
	Inspection	1	
	Total	23	20.53

Table 1 (continue)

	Attitude Topics	f	%
	Environment	11	
	Child Development	1	
	Living Creatures	1	
	Uncertainty	1	
	National Parks	1	
Other	Women's Employment	1	
	Health	1	
	Concrete Cultural Heritage	1	
	Strategic Planning Awareness Level	1	
	Extracurricular Activities (Parents)	1	
	Gender-Based Career Choices	1	
	Total	21	18.75
	Total	112	100.00

Having examined Table 1, it has been observed that attitude scale for various areas within education was developed. It has been determined that the most (36.61%) predominant one was the development of attitude scale for coursework. A large number of studies examining the relationship between attitudes towards coursework and the success in a given course, alongside the necessity of using an attitude scale about the courses in question, are considered to be what causes it. Moreover, having analysed the frequency values in the table, it is worth noting that there are multiple scale development studies focusing on just one subject (e.g. ten studies on science & technology courses, eleven studies on environment, etc.).

Number of Tested Items and Findings Related to the Size of Sample Group

Although there is no exact criterion in the literature on the number of items tested or the size of the sample group, some researchers state that the sample size should be at least five times the number of items tested, whilst other researchers suggest that it ought to be ten times the size of the sample (Child, 2006; Gorsuch, 1983; Kline, 1994, Tavşancıl, 2005). The number of items tested and the size of the group in which the trial was carried out in the studies considered in this context are given in Table 2.

Table 2

Distribution With Regards to the Number of Tested Items and the Size of Sample Group Ratio

Criterion	f	%
Less than fivefold	11	9.82
At least fivefold	50	44.64
Tenfold and above	46	41.07
No information	5	4.46
Total	112	100.00

Having examined Table 2, samples that are ten times the number of the tested items in forty-six of the items were studied. In addition, five times the number of the tested items were considered in fifty of the items, and the trial application was made on a group much less than five times the number of the testes items in eleven of the items. In five of the studies, there was information missing related to the number of items or number of people involved, and thus the findings of this study could not be reached.

Category Numbers of Rating Expressions

The results of the rating expressions used in the articles reviewed are presented in Table 3.

Table 3

Distribution Related to Rating Expressions Category Numbers

Number of Category	f	%
3-categories	7	6.25
4-categories	2	1.79
5-categories	97	86.61
No information available for category number	6	5.36
Total	112	100.00

Having examined Table 3, one observes that the general trend (86.61%) was to form the rating expressions in five categories. Tavşancıl (2005) states that ratings in the Likert type attitude scale can be either 3, 5, 7, 9, or even 11. However, in the literature, in parallel to the findings of this study, it has been found out that 5-rating expressions were generally preferred in those studies (Tavşancıl, 2001 Wiersma, 2000). In addition, having examined the studies preferring category 3, it has been seen that these studies were mostly oriented around primary school children, and three rating expressions were chosen because of their ability to facilitate understanding and response for small age groups.

Use of Terminology

As in other studies, some terminology is used interchangeably in attitude scale development studies. It is possible to come across terminology such as the survey-inventory-test, which is frequently used in attitude scales. Table 4 shows the differences in terminology used in this study.

Table 4

Distribution in relation to the Distribution of the Terminology Used

Term Used	f	%
Attitude scale	106	94.64
Attitude survey	6	5.36
Attitude inventory	-	-
Total	112	100.00

Having examined Table 4, it is possible to say that the attitude scale was generally used correctly (94.6%). In addition, in some studies, the scale and survey terminology were combined, and the survey was used instead of the scale.

Findings Related to Theoretical Knowledge and Operational Definitions

The first theoretical information based on the questioning of attitude scale development include information such as: "have the theoretical foundations, whose attitude scale was developed, been presented in detail in the study reviewed?", "has the operational definition been made with regards to this construct?", and, if yes, "has the operational definition been made correctly?" The frequency values obtained for these questions are given in Table 5.

Table 5

Distribution with Regards to the Theoretical Presentation of the Items Examined

Theoretical Foundation Presentation			Operational Definition				
	f	%		Correct	Partial	False	f	%
Yes	66	58.93	Yes	12	12	-	24	21.43
Partial	37	33.04						
No	9	8.04	No	88	-	-	88	78.57
Total	112	100.00	Total				112	100.00

Upon examining Table 5, it has been observed that in nine of the articles (8.04%) examined, the theoretical foundations of the structure to be measured were not included; in thirty-seven studies (33.04%), the information obtained was insufficient; and in sixty-six studies (58.93%), a sufficient amount of theoretical information was obtained. Having examined the studies where the theoretical information was incomplete or inadequate, it has been observed that it was mostly present in studies whose main purpose was not to develop an attitude scale. It has previously been noted that articles that do not focus on developing an attitude scale but nevertheless aim to deal with a relationship between attitude and other variables, as well as that develop an attitude scale as needed (and hence including these development stages in the report section) are also included in the study. In these studies, it was found that the attitude scale was developed and that the reporting of the psychological structure of this scale was sometimes overshadowed and not sufficiently included in the report.

Upon examining the section where the operational definition is questioned, it has been witnessed that there was no operational definition in majority of the studies (78.6%). Having examined a number of studies in which the definition of an operational definition is made, it has been found that twelve of the twenty-four studies were partially correct, whilst the general trend was to make the definition correctly.

Findings with Regards to Item Writing and Trial Application

One of the most important stages of scale development is item writing and trial application. In this context, the frequencies related to the articles reviewed within the

scope of the study with the criteria determined within the framework of item writing and trial application is presented in Table 6.

Table 6

Frequencies of Item Writing and Trial Application

	Y	P	N	NI
	f (%)	f (%)	f (%)	f (%)
Compliance of the item with the principles of item writing	29 (25.89)	43 (38.39)	8 (7.14)	32 (28.57)
Distribution of positive/negative items	38 (33.93)	34 (30.36)	12 (10.71)	28 (25.00)
Distribution of cognitive, affective, and dynamic components of attitude	25 (22.32)	42 (37.50)	16 (14.29)	29 (25.89)
Compliance between the rating expressions used and written items	69 (61.61)	6 (5.36)	8 (7.14)	29 (25.89)
Expert opinion on prepared items	37 (33.04)	58 (51.79)	17 (15.18)	-

Y: Yes Suitable, P: Partial Suitable, N: Not Suitable, NI: No Information

Upon looking at Table 6, according to the results obtained from the examination of the compliance between the items and the principles of item writing, twenty-nine of the articles (25.89%) took into account the principles of the item, forty-three of them (38.39%) were partially suitable, eight of them (7.14%) was definitely not suitable, and thirty-two of them (28.57%) did not contain any information about any of the items. In addition, in the scope of the articles reviewed, a sample section of the items that did not comply with the principles of item writing was given below.

- … make me feel both uneasy and confused. (Two judgments)
- Develops the ability to analyse, synthesize, and interpret … (Multiple judgments)
- … are highly exaggerated: there is already a large number in nature; it does not matter whether or not a few of them disappear.
- I watch TV and listen to radio programs related to …
- I feel fear and excitement before … exams.
- It enables me to display my knowledge and capacity in …
- I know … (i.e. keyboard, screen, mouse, printer, scanner, floppy disk, CD-ROM, disc, etc.) and their functions.
- I have knowledge about the … and …
- Desertification does not take place in … (Factual)
- … is a serious environmental problem. (Factual)
- Although …, they have an important place in nature, and therefore I am against them being killed.
- I hate most …; however, I don't kill them. (Contains frequency phrases)

To avoid any ethical violation while reporting the unsuitable items, the attitude objects in the items are left blank. In the articles reviewed, it is possible to find more

items besides the above in cases which are contrary to the principles of item writing. Another remarkable point is the inconsistency of the attitude objects and items in the study. According to the sources of Likert type attitude scale development (Tavşancıl, 2005; Tezbaşaran, 1996; Turgut, 1977), the distribution of positive/negative expressions in attitude scale items should be very close to or equal to each other. In this context, it is possible to say that the distribution was balanced in thirty-eight of the articles (33.93%) reviewed. However, thirty-four of them (30.36%) were partially followed, and twelve of them (10.71) did not have a balanced distribution. Twenty-eight of the articles (25.89%) did not provide any information on positive/negative items. It is possible to say that the balance of the distribution was paid attention to in most of the articles reviewed, it was considered in only part in a small number of the articles. Another criterion that should be taken into consideration is the distribution of attitude with regards to cognitive, affective, and dynamic components. The distribution was balanced in twenty-five of the articles (22.32%) covered in the study. In sixteen of them (14.29%), the distribution was not balanced. However, it should be kept in mind that the three components of the attitude have dimensions that should be considered while measuring the attitude.

After examining the compliance between the rating expressions and the items in the articles reviewed, it is possible to say that sixty-nine of them (61.61%) were suitable, six (5.36%) were suitable only in part, eight (7.14%) were not suitable, and there is no information in twenty-nine of them (25.89%). It can be said that one of the most carefully cited criteria in item writing and trial application is the consistency of items and rating expressions.

Finally, no expert opinion has been received in seventeen articles (15.18%) in the criterion of submitting the prepared items to the expert opinion, it has been observed that expert opinion has been partially received in fifty-eight of the articles (51.79%), and adequately received in thirty-seven of them (33.04%). The reason for this is that in some studies only the subject area expert's opinion is taken. In others, the opinion of all Turkish language, measurement, and evaluation experts are taken.

Reliability Findings

The reliability studies conducted in the articles have been examined under headings such as have there been any studies conducted on the degree of reliability of the final scale? (e.g. the Cronbach's Alpha, test-retest, etc.), and are the reliability studies conducted adequate? The frequencies related to these titles are presented in Table 7.

Table 7

Distribution With Regards to Reliability in the Articles Reviewed

	Y	P	N	NI	Total
	f (%)	**f (%)**	**f (%)**	**f (%)**	**f (%)**
Has the reliability study been carried out?	108 (96.42)	-	4 (3.57)	-	112 (100)
Are the reliability studies conducted adequate?	103 (91.96)	5 (4.46)	-	4 (3.57)	112 (100)

Y: Yes Suitable, P: Partial Suitable, N: Not Suitable, NI: No Information

Reliability studies were conducted for one hundred and eight of the articles (96.42%) reviewed, but not for four (3.57%). While determining the adequacy of reliability studies, attention was paid to the points such as the suitability of the reliability evidence used for the scale, its correct reporting, its correct interpretation etc. As for the adequacy of the reliability studies, one hundred and three (91.96%) were found to be adequate by the researchers, five (4.46%) were found to be partially adequate, and no reliability study was conducted for four of them and thus there is no information thereof. However, the frequencies obtained when the techniques used for proof of reliability were examined are presented in Table 8.

Table 8

Distribution with Regards to the Reliability Determination Methods Used in the Articles

Methods	f	%
Cronbach Alpha	108	96.43
Item/Total Test Correlation	68	60.71
Test/Retest	16	14.29
Split-half reliability	3	2.68

Note. More than one reliabity methods were used in some studies.

Having examined the reliability determination methods used in studies, it has been found out that the Cronbach Alpha internal consistency coefficient was used in %96.43 of the studies, which was to be expected given that this coefficient is used to determine reliability in multiple-rated items (i.e. the Likert-type attitude scale) (Crocker and Algina, 1986, Erkuş, 2003). In addition, other reliability determination techniques were also used in other studies. Item total test correlation was considered in sixty-eight of these studies (60.71%), the test-retest method in sixteen (14.29%), and the split-half reliability in three of them (2.68%).

Findings Related to Validity

It has been questioned in the studies reviewed as to whether or not a validity study was carried out. The adequacy of the validity determination methods used was evaluated.

Table 9

Distribution With Regards to the Validity Studies

		Yes			No	Total
		Suitable	**Partially Suitable**	**Not Suitable**	**No**	**Total**
Has the validity study been carried out?	f	61	27	10	14	112
	%	54.46	24.11	8.93	12.50	100.00

Upon examined Table 9, it can be seen that either validity determination study was not applied in 14 of 112 studies (12.50%), or no report was provided whatsoever. Taking into account that validity is an indispensable feature of a measurement instrument and, provided that a new instrument is being developed, it must be proved that it is a valid instrument for it to be used. When the appropriateness of validity methods used in the remaining 98 studies are examined, it has been seen that the validity studies of only some of them (54.46%) were adequate, and moreover that even if the proof of validity was presented, the remaining studies were either incomplete (24.11%) or incorrect (8.93%). These studies were examined in detail in order to describe these inaccuracies. It has also been discovered that the incompleteness or inaccuracies in the validity evidence leading to ticking suitable or not suitable were due either to the mistakes made in the selection of validity method, or to a mistake in the implementation or reporting of the chosen method.

As a next step, studies in which validity studies were conducted were examined in detail, and the frequency information was obtained regarding which validity methods were used. If more than one proof of validity was presented in the article, all of the methods were therefore counted. The information is given in Table 10.

Table 10

Distribution with Regards to Validity Methods

Methods	f	%*
Exploratory Factor Analysis	87	77.68
Confirmatory Factor Analysis	37	33.04
Criterion Validity	14	12.50
Expert Opinion	12	10.71
Item Discrimination	11	9.82

*More than one reliability methods were used in some studies.

Upon closer inspection of Table 10, it has been seen that a total of five different methods were used as a validity determination method. The most commonly used of these methods (77.68%) was exploratory factor analysis, which is a proof of construct validity. Having examined the articles included in the study, it has been observed that the factor analysis study was applied as a basic method in the validity study, and that the other studies were applied to provide additional evidence beyond the factor

analysis findings. It has been observed that the relationship between the obtained attitude scores and another criterion was considered in fourteen articles (12.50%), the presenting evidence for construct validity was considered in twelve articles (10.71%), and the significance of the difference between the attitude scores of the lower and upper groups was tested in eleven studies (9.82%). Furthermore, in addition to the exploratory factor analysis, confirmatory factor analysis was conducted in more recent studies. In addition, in the majority of studies (except for three), the confirmatory factor analysis study and the exploratory factor analysis study were conducted on the same data set.

The studies using factor analysis as a proof of validity were broadly evaluated in terminology of whether the KMO and Barlett test results were reported or not, and in terminology of the adequacy of the reporting of the factor analysis studies. The results are presented in Table 11.

Table 11

Distribution With Regards to the Reporting of Factor Analysis Studies

	Yes	Partially	No	Total
	f (%)	f (%)	f (%)	f (%)
Availability of KMO and Barlett Report	60 (68.9)	2 (2.3)	25 (28.7)	87 (100)
Suitability of FA Report	54 (62.1)	24 (27.6)	9 (10.3)	87 (100)

In taking a look at Table 11, one notices that the KMO and Barlett test results were reported adequately in sixty of the eighty-seven studies (68.96%), while they were never mentioned in twenty-five of them (28.74%), and given that the direct factor analysis results were included. However, it is appropriate to present the results of factor analysis after proving that the data is suitable for factor analysis. Furthermore, it has been remarkable that, in some studies, the factor analysis data was quite small, leading to the uncertainty of whether or not the factor analysis was appropriate for implementation. In two studies where partially was selected, the KMO value was given and compared with the accepted criteria in literature, and its suitability was evaluated accordingly, even though there was no mention of either Barlett's value.

When the analysis of the factor analysis results is examined, as seen in Table 11, only fifty-four of the eighty-seven studies (62.06%) were conducted in a proper manner, and the remaining studies were incomplete (27.59%) or inaccurate (10.34%).

In fourteen studies using the criterion-related validity study, the suitability of the criterion used and whether the psychometric characteristics of the criterion were reported were examined. The results are presented in Table 12.

Table 12

Distribution With Regards to the Suitability of the Criterion Used

	Yes	Partially	No
Suitability of the criterion	12	2	-
Reporting the psychometric properties of the criterion	12	2	-

Upon examining Table 12, it has been seen that the psychometric properties of the criterion used in almost all criterion validity studies were presented, and an appropriate criterion was used. The suitability of the criterion was evaluated both in the sense that it measures the same or similar structures with the developed attitude scale, as well as in the sense that it has acceptable validity reliability coefficients. It is only in two studies that only the reliability proof of the criterion was presented, and that there was no mention of validity proof. After detecting important inaccuracies in scale development studies in literature, the importance of performing a rigorous research process about the suitability of the criterion, as well as about reporting and sharing this process with the reader has been realized. In the articles reviewed, it has been found that these studies were conducted accordingly.

Finally, the articles were examined in order to determine whether the test statistics for the score distribution of the final scale were included. In forty-eight studies, information was provided about the distribution of scores, while sixty-four studies did not provide such information. When examining the articles, it has been seen that the main purpose of the studies that reported such information was not to develop scale. Hence, it has been concluded that this process was conducted because it was necessary for other analyses performed on the scale however not because it was seen as a necessity of scale development study.

Discussion, Conclusion and Suggestions

The attitude topics were examined in studies reviewed within the scope of this study and were divided into 6 main headings: attitude scales for coursework in general, attitude scales for the profession, attitude scales for technology, attitude scales for a teaching method-technique, attitude scales for other education-related items, and attitude scales for other topics. Having examined the sub-study topics and frequency values of these subjects in greater detail, it has been found out that the most studied area was developing an attitude scale for coursework. In addition, it has been concluded that there were multiple attitude scale development studies on the same subject. There are several reasons for this: for example, the researcher might not have been able to use the existing studies due to the missing or faulty parts in the existing attitude scale development studies in the literature or in the reports of the researchers. This is interpreted as an indicator of a serious loss of labour and time when considering the difficulty of scale development process. Moreover, the researcher might not have conducted the literature review sufficiently detailed, or might not have noticed the attitude scales on the topic they wished to study. In some cases, there may be situations where existing scales measure constructs close to each other but do not

measure the same constructs. However, as a result of the research conducted, it has been found that the samples of this situation were quite low.

In the attitude scales developed, the preferred category numbers for rating expressions were examined. It has been found out that the number of the most preferred category number was 5 in accordance with the literature. The next choice is 3. However, it has been concluded that the 3-point rating was often preferred for small age groups. Having examined the terminology used in the attitude scales developed (attitude scale, attitude survey, attitude inventory), it is possible to say that the terminology was generally used correctly. In some studies, it has been seen that attitude scale was used instead of attitude survey. In this context, it is possible to say that the researchers did not have thorough knowledge about what the scale-survey-inventory-test terminology meant and where they were used.

After examining the numbers preferred in the sample sizes in the articles, it has been observed that those who preferred ten times or above the number of items chose the adequate size of the sample, while the majority of the studies examined factor analysis on the sample less than ten times the number of items. Certain studies in the literature have shown that the results of factor analysis in samples less than ten times the number of items has been found to be inaccurate. It has been emphasized that much larger samples should be used (Kline, 2013).

In the attitude scale development studies, it has been concluded that, although the information on the theoretical basis of the structure to be measured was generally presented, the operational definition of the object of attitude in the study was mostly not made, and was missing. In certain parts of the attitude scale development studies, it has been observed that the criteria to be met in the principles of item writing were ignored, factual statements were given, more than one judgment was included in each attitude item, the written item and the attitude object were unrelated, and that the statements that reported frequency were included. It is possible to say that this situation is contrary to the scale development process. Each of the stages and requirements for developing scales is stated in all published sources (Cohen and Swerdlik, 2010; De Vellis, 2003; Erkuş, 2012; Murphy and Davidshofer, 2005). The most important step that should not be skipped is the adequate operational definition of the concept to be measured. In some attitude scales, it is possible to say that the principles of item writing were followed correctly, and that all other stages (explanation of conceptual and operational definition, reliability and validity studies) were suitable. Hence, it can be said that each of these stages is related to each other and that the criteria, which were either overlooked or not included in one unit, can affect the other stages.

Reporting reliability evidence is a must have in a scale development study. Having examined the studies, it is possible to say that reliability was generally included and was done correctly. But on the other hand, a significant number of published studies appeared to show no evidence of validity of the scale. It has been concluded that the most widely used validity determination method was factor

analysis, and even provided to be a basis for this study. Other validity determination methods were found to be the second or third preferred methods when it was necessary to present more than one piece of evidence. It has been found out that, in the factor analysis studies, the KMO and Barlett statistical results were reported in general, while some of these values were never included in the report or were reported to be incomplete. It has been concluded that the criterion used in all of the studies where the criterion validity determination method used was a suitable criterion, and that the psychometric properties of the criterion were included in the report.

In recent studies, it has been seen that the confirmatory factor analysis was performed besides exploratory factor analysis. In addition, it has been observed in the majority of studies (except for three) that the confirmatory factor analysis study and the exploratory factor analysis study were conducted on the same data set. In the literature, in order to determine the measured structure, it is stated that exploratory factor analysis should be done in order to determine the structure and then continue with confirmatory factor analysis on a different sample (Henson and Roberts, 2006; Worthington and Whittaker, 2006).

Attitude scale development studies within the scope of this study have attempted to determine the points that are often inaccurate or incomplete in the literature. It is recommended that scale developers consider this process as an important and rigorous process, and that they meticulously consider the points underlined in the study. Within the scope of this study, studies that develop an attitude scale were examined; similar scale adaptation studies should also be examined. In addition, only the domestic literature has been examined: studies that examine international could also be investigated. The results of the study suggest that there are a lot of inaccurate and incomplete attitude scales in the current literature. In addition, it is worth noting that there at times may be more than one scale developed on the same topic. Both of these issues mean a serious loss of labour and time for researchers. It is believed that there is a need for a test centre at the national level that both supervises and coordinates the scales used in order to solve this problem.

References

Acar-Güvendir, M. and Özer-Özkan, Y. (2015). The examination of scale development and scale adaptation articles published in Turkish academic journals on education. *Electronic Journal of Social Science*, *14*(52), 23-33. doi: 10.17755/esosder.54872

Anastasi, A. (1988). *Psychological testing* (6th Ed.). New York, NY: MacMillan Publishing Co. Inc.

Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması.* Ankara: ÖSYM Yayınları.

Bilgin, N. (2006). *Sosyal bilimlerde içerik analizi.* Ankara: Siyasal Kitabevi.

Boztunç-Öztürk, N., Eroğlu, G. and Kelecioğlu, H. (2014). A review of articles concerning scale adaptation in the field of education. *Education and Science*, *40*(178), 123-137. doi: 10.15390/EB.2015.4091

Child, D. (2006). *The essentials of factor analysis.* London: A&C Black.

Coaley, K. (2010). *Psychological assessment and psychometrics.* California, CA: Sage Publications.

Cohen R. J., and Swerdlik M. E. (2010). *Psychological testing and assessment*. Boston, MA: McGraw-Hill Companies.

Crocker, L., and Algina, J. (1986). *Introduction to classical and modern test theory*. USA: Rinehart and Winston Inc.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York, NY: Harper Collins Publishers.

Çüm, S., and Koç, N. (2013). The review of scale development and adaptation studies which have been published in psychology and education journals in Turkey. *Journal of Educational Sciences and Practices, 12*(24), 115-135.

De Vellis, R. F. (2003). *Scale development theory and applications.* California, CA: SAGE Publication Inc.

Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: Türk Psikologlar Derneği Yayınları.

Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar. *Turkish Journal of Psychology, 13*(40), 17-25.

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme-1: Temel kavramlar ve işlemler.* Ankara: Pegem Akademi.

Gorsuch, R. L. (1983). *Factor analysis.* New Jersey, NJ: Lawrence Erlbaum Associates.

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

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, *21*(5), 967-988. doi: 10.1177/014920639502100509

Kaya-Uyanık, G., Güler, N., Taşdelen-Teker, G. and Demir S. (2017). Investigation of scale development studies conducted in educational sciences published in Turkey by many-faceted rasch model. *Journal of Measurement and Evaluation in Education and Psychology*, *8*(2), 183-199. doi: 10.21031/epod.291367

Kilmen, S. (2017). Ölçme ve değerlendirmede temel kavramlar. In R. N. Demirtaşlı (Ed.), *Eğitimde ölçme ve değerlendirme* (pp. 25-56). Ankara: Anı Yayıncılık.

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

Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher and C. Schatsschneider (Eds.), *Applied quantitative analysis in the social sciences* (pp. 171-207). New York, NY: Routledge.

Likert, R. (1932). *A technique fort he measurement of attitudes*. New York: Archives of Psychology.

Miles, M., and Huberman, M. A. (1994). *An expanded sourcebook qualitative data analysis.* London: Sage Publications.

Murphy K. R., and Davidshofer C. O. (2005). *Psychological testing: Principles and applications*. New Jersey, NJ: Pearson Education International.

Özgüven, İ. E. (2011). *Psikolojik testler.* Ankara: Pegem Yayınları.

Rust J., and Golombok S. (1997). *Modern psychometrics: The science of psychological assessment.* New York, NY: Routledge.

Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., and Lankau, M. J. (1993). Improving construct measurement in management research: comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type ınstruments. *Journal of Management, 19*(2), 385-417. doi: 10.1016/0149-2063(93)90058-U

Schultz, D. P., and Schultz, S. E. (2007). *Modern psikoloji* tarihi [A history of modern psychology]. (Y. Aslay, Trans.). İstanbul: Kaknüs Yayınları. (Orijinal kitabın yayım tarihi 2004)

Tavşancıl, E. (2005). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel Yayıncılık.

Tavşancıl, E., and Aslan, E. (2001). *Sözel, yazılı ve diğer materyaller için içerik analizi ve uygulama örnekleri.* İstanbul: Epsilon Yayınları.

Tavşancıl, E., Güler, G., and Ayan, C. (2014, June). *Review of attitude scales developed in Turkey between 2002 and 2012 regarding scale developing process.* Paper presented at the Annual Meeting of the 4th Congress on Measurement and Evaluation in Education and Psychology, Hacettepe University, Ankara, Turkey.

Tezbaşaran, A. A. (2008.) *Likert tipi ölçek hazırlama kılavuzu*. Ankara: Türk Psikologlar Derneği Yayınları.

Tezbaşaran, A. A. (1996). *Likert tipi ölçek geliştirme kılavuzu*. Ankara: Türk Psikologlar Derneği Yayınları.

Thurstone, L. L. (1929). Theory of attitude measurement. *Psychological Review*, *36*(3), 222-241. doi: 10.1037/h0070922

Turgut, F., and Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi Yayıncılık.

Turgut, M. F. (1977). *Ölçmede geçerlik* [Yayınlanmamış ders notları 3].

Wiersma, W. (2000). *Research methods in education: an ıntroduction.* Needham Heights, MA: Allyn and Bacon, A Pearson Education Company.

Worthington, R. L., and Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806-838. doi: 10.1177/0011000006288127

The Ethical Committee Approval

The ethical committee approval is not compulsory for this research because it was sent to our journal before 01.01.2020.

Türkiye'de 2002-2018 Yılları Arasında Geliştirilen Tutum Ölçeklerinin Ölçek Geliştirme Süreci Açısından İncelenmesi

Gül Güler [1]
İstanbul Aydın Üniversitesi

Cansu Ayan [2]
Ankara Üniversitesi

Öz

Bu araştırma kapsamında belirli yıllar içerisinde eğitim alanında kullanılan tutum ölçeği geliştirme çalışmalarının ölçek geliştirme süreçlerine uygunluğunun incelenmesi amaçlanmaktadır. Bu sayede ölçek geliştirme süreci ile ilgili alanyazında sıklıkla karşılaşılan sorunlar belirlenmiş olacaktır. Ayrıca bu çalışma bundan sonraki tutum ölçeği geliştirme çalışmaları için de yol gösterici ve bilgi verici olması açısından önemli görülmektedir. Bu kapsamda ULAKBİM'de taranan yerel dergiler ve elektronik ortamda tam metinlerine ulaşılabilen eğitim dergilerinde tarama yapılmış ve 2002-2018 yılları arasında Türkiye'de eğitim alanında yapılmış tutum ölçeği geliştirme çalışmaları, tutum ölçeği geliştirilirken dikkat edilmesi gereken noktalar dikkate alınarak incelenmiştir. Bu çalışma nitel bir çalışma olup, araştırmacılar tarafından önceden belirlenen bir kontrol listesi kullanılarak, her bir ölçüt (kriter) için uygun olan ve olmayan çalışmaların frekansları belirlenmiştir. Çalışma sonucunda tutum ölçeği geliştirme çalışmalarında, alanyazında sıklıkla yanlış yapılan ya da eksik bırakılan noktalar saptanmıştır. Test ve ölçek geliştiricilere, bu sürecin önemli ve titizlik gerektiren bir süreç olduğunu unutmayıp çalışma kapsamında altı çizilen noktalara duyarlı biçimde yaklaşmaları önerilmiştir.

Anahtar sözcükler: Tutum, tutum ölçeği, ölçek geliştirme.

Etik Kurul Kararı: Bu araştırma, dergimize 01.01.2020 tarihinden önce gönderildiği için etik kurul kararı zorunluluğu taşımamaktadır.

[1] *Sorumlu Yazar*: Dr. Öğr. Üyesi, Eğitim Fakültesi, Temel Eğitim Bölümü, Sınıf Eğitimi Anabilim Dalı, E-posta: gulyuce@aydin.edu.tr, https://orcid.org/0000-0001-8626-4901
[2] Arş. Gör. Dr., Eğitim Bilimleri Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, E-posta: cayan@ankara.edu.tr, https://orcid.org/0000-0002-0773-5486

Amaç ve Önem

Alanyazında ölçek geliştiren çalışmalar incelendiğinde hem sayıların her geçen yıl ciddi miktarda arttığı hem de var olan çalışmalarda ciddi teknik sıkıntılar olduğu görülmektedir. Alanyazında yer alan yanlış uygulamalar, kötü örnek oluşturmakta ve bazı sorunların süreğen (kronik) duruma gelerek diğer çalışmalarda da aynı biçimde tekrarlanmasına neden olmaktadır. Bu sorunların ve eksiklerin belirlenmesi daha sonra yapılacak çalışmalara da ışık tutulması anlamında önemli görülmektedir. Bu araştırma, belirli yıllar içerisinde yapılan tutum ölçeği geliştirme çalışmalarının incelenmesi ve tutum ölçeği geliştirme sürecinin ne ölçüde dikkate alındığının bir bütün olarak sunulması, alanyazında var olan sorunları betimlemesi ve bundan sonraki tutum ölçeği geliştirme çalışmaları için de bilgi verici olması açısından önemli görülmektedir. Bu kapsamda, çalışmada 2002-2018 yılları arasında Türkiye'de eğitim alanında yapılmış olan tutum ölçeği geliştirme çalışmalarının ölçek geliştirme adımlarına uygunluğunun incelenmesi amaçlanmaktadır.

Yöntem

Bu araştırma, tutum ölçeği geliştiren çalışmalar, tutum ölçeği geliştirilirken dikkat edilmesi gereken noktalar dikkate alınarak incelendiğinden nitel bir araştırmadır. Çalışmada kapsamında doküman analizi yapılmıştır. Önceden belirlenen ölçütlere göre makaleler incelenmiş, her bir ölçüt için frekans değerleri elde edilmiştir. Bu kapsamda ilk adım olarak ulusal bilgi ve belge erişim hizmetleri sunmak amacıyla TÜBİTAK tarafından kurulmuş bir enstitü olan ULAKBİM'de taranan yerel dergiler ve elektronik ortamda tam metinlerine ulaşılabilen eğitim dergilerinde tarama yapılmış ve 112 makaleye ulaşılmıştır. Bununla birlikte, bu araştırma 01.01.2020 tarihinden önce yapıldığı için etik kurul kararı zorunluluğu taşımamaktadır.

Araştırmanın verileri üç temel bölümden oluşan bir kodlama listesi kullanılarak elde edilmiştir. Bu üç bölüm; yapılan çalışma ile ilgili ön bilgilerin sorgulandığı başlangıç bölümü, dört ana başlıkla ilgili izlenmesi gereken adımları sorgulayan maddelerden oluşmuş ve yanıtı *evet*, *hayır*, *kısmen*, *bilgi yok* biçiminde derecelendirilmiş ikinci bölüm ve son olarak çalışmalarda kullanılan geçerlik ve güvenirlik kanıtlarının neler olduğunu sorgulayan son kısımdan oluşmaktadır. Kodlama listesinin ikinci bölümü ayrıntılı olarak incelendiğinde kuramsal ve işevuruk tanımlar, madde yazımı ve deneme uygulaması, güvenirlik, geçerlik olmak üzere 4 ana başlıktan oluşmaktadır. Her bölümde bu bölüm içinde yapılması gereken noktaları sorgulayan ilgili maddeler yer almaktadır.

Veriler nitel araştırmalarda kullanılan bir analiz türü olan içerik analizine tabi tutulmuştur. İçerik analizinin bir türü olan kategorisel analiz yöntemi uygulanmış ve her bir kategoriye ait frekanslar hesaplanmıştır. Son olarak her noktayla ilgili olarak frekans değerleri hesaplanmış ve raporlaştırılmıştır.

Bulgular

Bu araştırma kapsamında incelenen çalışmalarda ele alınan tutum konuları incelenmiş ve genel olarak bir derse yönelik tutum ölçekleri, mesleğe yönelik tutum

ölçekleri, teknolojiye yönelik tutum ölçekleri, bir öğretim yöntem-tekniğine yönelik tutum ölçekleri, eğitimle ilgili diğer öğelere yönelik tutum ölçekleri ve diğer konulara yönelik tutum ölçekleri olmak üzere altı temel başlığa ayrılmıştır.

İncelenen tutum ölçeği geliştirme çalışmalarında, genel olarak ölçülmesi hedeflenen yapının, kuramsal temellerine dair bilgiler sunulmuş olmasına karşın, çalışmadaki tutum nesnesinin işe vuruk tanımının büyük çoğunlukla yapılmadığı, eksik bırakıldığı sonucuna ulaşılmıştır. İşe vuruk tanımı yapan çalışmalarda ise bu tanımlamanın genel olarak doğru biçimde yapıldığı belirlenmiştir. Tutum ölçeği geliştirme çalışmalarının bir bölümünde madde yazımı ilkelerinde uyulması gereken ölçütlerin göz ardı edildiği, olgusal ifadelere yer verildiği, her bir tutum maddesinde birden çok yargının yer aldığı, tutum objesi ile yazılan maddenin ilişkisiz olduğu ve sıklık bildiren ifadelere yer verildiği gözlenmiştir. Bu durumun ölçek geliştirme sürecine aykırı olduğunu söylemek olanaklıdır.

Son yıllarda yapılan çalışmalarda açımlayıcı faktör analizinin yanında doğrulayıcı faktör analizi çalışmasının da yapıldığı görülmektedir. Bunun yanında çalışmaların büyük çoğunluğunda (üçü hariç) doğrulayıcı faktör analizi çalışması ile açımlayıcı faktör analizi çalışmasının aynı veri seti üzerinde yürütüldüğü görülmektedir. Alanyazında ise, ölçülen yapının belirlenmesi için analizlerde öncelikle açımlayıcı faktör analizi yapılıp yapının belirlenmesi ve sonrasında farklı bir örneklem üzerinde doğrulayıcı faktör analizi ile devam edilmesi gerektiği belirtilmektedir (Henson ve Roberts, 2006; Worthington ve Whittaker, 2006).

Tutum ölçeği geliştirme çalışmalarında bir diğer ele alınması gereken güvenirlik çalışmalarıdır. Yapılan çalışmalar incelendiğinde, genel olarak güvenirlik çalışmalarına yer verildiği ve doğru bir biçimde yapıldığını söylemek olanaklıdır. Yayınlanmış çalışmaların önemli bir kısmında ölçeğe ilişkin herhangi bir geçerlik kanıtı sunulmadığı belirlenmiştir.

Tartışma, Sonuç ve Öneriler

Bu araştırma kapsamında tutum ölçeği geliştirme çalışmalarında, alanyazında sıklıkla yanlış yapılan ya da eksik bırakılan noktalar belirlenmeye çalışılmıştır. Test ve ölçek geliştiricilerin, bu sürecin önemli ve titizlik gerektiren bir süreç olduğunu unutmayıp, çalışma kapsamında vurgulanan noktalara duyarlı biçimde yaklaşmaları önerilmektedir. Bu çalışma kapsamında bir tutum ölçeği geliştiren çalışmalar incelenmiştir, benzer biçimde ölçek uyarlama çalışmaları da incelenebilir.

Etik Kurul Kararı

Bu araştırma, dergimize 01.01.2020 tarihinden önce gönderildiği için etik kurul kararı zorunluluğu taşımamaktadır.