Education and Science

Vol 40 (2015) No 178 85-102

Thematic Content Analysis of Scale Development Studies Published in the Field of Science and Mathematics Education

Şeyda Gül [1], Mustafa Sözbilir [2]

Abstract

The aim of this study is to analyse the scale development studies published by researchers in Turkey in terms of content and methodological aspects. For this aim, the papers published in six major academic journals in the field of educational sciences during the years 2000-2013 were reviewed in accordance with criteria of inclusion of the research and a total of 22 articles were included in this study. The analysis was carried out through a meta-synthesis. The findings indicated that the majority of the studies focused on attitudes and the scales were mostly developed in the field of mathematics education. It was also found that construct validity was generally achieved and confirmatory factor analysis was limited as validity analysis and the analysis towards methods of exploratory factor analysis were generally conducted in moderate level. The findings also indicated that the most preferred reliability analysis method was the internal consistency. Finally, various recommendations were given in accordance with the findings obtained.

Keywords

Scale development

Thematic content analysis/Meta-synthesis

Science and mathematics education

Article Info

Received: 10.19.2014

Accepted: 03.06.2015

Online Published: 05.04.2015

DOI: 10.15390/EB.2015.4070

Introduction

In our changing world, the needs of society are rapidly changing and scientific research about education is important for training the man power to meet those needs (Çiltaş, Güler & Sözbilir, 2012). Scientific research whose results affect the policies and practices in all areas of science form an empirical basis for applications; and they can be also used as a guide for practitioners in professional activities. In addition, the results of scientific research in the field of education are important content resources in terms of the studies published in the form of textbooks, encyclopedias and manuals (Karadağ, 2009).

Scientific research has generally social purposes. Research is an effort to and a function of knowledge production which tries to recognize the community, to describe its profile, to find out relationships related to variables and answers to questions such as why, how, when, and where, related to factors that lead to arise a problem (Özdamar, 2003, p:2). Moreover, scientific research can be done only through data (Ercan & Kan, 2004). In a scientific research, the scales have an important role in obtaining, digitizating and recording the data and in putting these data in application (Karagöz & Ekici, 2004).

[1] Ataturk University, Kazım Karabekir Faculty of Education, Department of Biology Education, Turkey, seydagul@atauni.edu.tr
[2] Ataturk University, Kazım Karabekir Faculty of Education, Department of Chemistry Education, Turkey, sozbilir@atauni.edu.tr

In addition to showing mathematical properties of measurement results in substance, the concept of scale is used in order to collect information from the target person or persons, the system, subject or content in the field of behavioral sciences such as education and psychology (Yurdugül, 2005). Of course, there are some characteristics which should be taken into account in selection of scales/measurement tools used in this process of information collection. The most important ones of these characteristics are compatible with research subject, the reliability and validity of measurement tool as conducting validity and reliability studies can provide suitable data for the scale developers and practitioners in terms of accuracy of the research results in a scientific research (Ercan & Kan, 2004). Although there are different classifications with regard to validity and reliability measures in the literature (Fraenkel & Wallen, 2000; McMillan & Schumacher, 2009; Topu, Baydaş, Turan & Göktaş, 2013), validity and reliability measures proposed in scale development studies can be generally summarized as follows:

Validity Measures;

- To get expert opinion
- To give information about the data collection tool
- To give information about the data collection process
- To disclosure who/whom to be done data collection process and evaluation by
- To describe the assumptions and limitations
- To explain the properties of the sample
- To include the participants who are volunteer
- To describe application process of research
- Item number

Reliability Measures;

- To get expert opinion
- To disclosure the methods of reliability calculations
- To do reliability study among evaluators
- To get participant opinions (control)
- To control the consistency among the data
- Item number

In addition to the above validity and reliability measures, a wide variety of methods and techniques can be used in the process of scale development. Nevertheless, the development of measurement tools can be handled in two stages called as the design of instrument and pilot testing. There are different concepts of validity that are specific to each stage. While content and face validity come into prominence in the stage of designing the measasurement tool, factorial, criterion (yordama, uygunluk) and construct (convergence and divergence) validities are taken into account in the application stage (Yurdugül & Bayrak, 2012).

In overall evaluation, scale development studies are long-term studies and the researchers who will develop the scale should have the field information related to variable to be measured, scale development and statistical knowledge, and skills in sufficient level. Besides, it is an important issue to do necessary changes over time in parallel with developments related to properties examined in scales and to re-examine the pometric properties of measurement tool in different samples and time (Azaltun, 2008). In literature, there were studies that gave information to researchers about methods and applications used in scale and scale development and consolidating the process of scale development (Azaltun, 2008; DiStefano, Zhu & Mindrila, 2009; Ercan & Kan, 2004; Erkuş, 2012; Yurdugül, 2005). Nevertheless, it is suggested that there is, especially in recent years, a remarkable increase in the scale development studies in Turkey, and also many scholars from the measurement and evaluation area or close/remote areas, due to the driving force of the publication compulsion on them, think unfortunately scale development or scale adaptation studies as "the easiest" way of overcoming this obstacle and this situation causes many errors (Erkuş, 2007). In addition, it seems that

similar or different processes were applied in these studies carried out on certain issues independently and different results were obtained. Especially, as examining the studies in the field of science, science-technology and mathematics education, it seems that there are a lot of studies focused on the topics of attitudes, self-efficacy, anxiety and others (perception, understanding, belief etc.) in general. Moreover, being generally evaluated, it seems that these scale development studies obtained the findings on finding out various dimensions with individual efforts of researchers. Therefore, it is important to do reliable quality top studies for interpreting this mass of information and for leading to new studies (Akgöz, Ercan & Kan, 2004). The fact that the most used data collection tool is scale in the studies examining articles and master or doctoral theses in this area makes the stages to be considered in scale development more important (Acar-Güvendir & Özer-Özkan, 2015). Therefore, that the concept of scale is accurately perceived by especially researchers can only be accomplished through knowing how the research was implemented and what kind of results were found. The most important steps to be taken in order to accomplish this are probably to determine the stages in scale development process, to find out present situation and shortcomings and thus to examine the scale development studies in a more detailed way.

As examining the previous studies related to this topic in other countries, it is seen that these studies which are limited in number focused mostly on a single journal, made analyses with limited subjects and statistical techniques and the findings were mostly interpreted with quantitative techniques (Dawis, 1987; Fassinger, 1987; Kahn, 2006; Tinsley & Tinsley, 1987; Worthington & Whittaker, 2006). For instance, it is seen that the studies such as Tinsley and Tinsley (1987) and Kahn (2006) focus on the factor analysis of publicaitons only in a journal and Martens (2005) and Quintana & Maxwell (1999) offer a general assessment focusing on the structural equation model in the same journal. However, Çüm (2013)' study on the examination of the papers published in the field of psychology and educational sciences, Mor-Dirlik (2014)'s study on the examination of 5 thesis in the field of educational sciences, and Acar-Güvendir ve Özer-Özkan (2015)'s study generally examined the similarity and differences of scale development/adaptation studies in the field of education in selected three journals with survey method. In addition, there is a study which is a descriptive presentation of totally 62 papers including developed/adapted scales/achievement tests in the field of science education with document analysis by Tosun and Taşkesenligil (2014). However, there is no study which examines the whole processes of validity and reliability in the process of scale development in detail and synthesizes present situation based on the scale development studies in framework of the main themes created in this research in Turkey. Therefore, it can be said that there is a need for doing a research of thematic content analysis (meta-synthesis) in this subject. Such a study is considered a guide for the reserchers for being aware of present scale development studies and for doing similar research in the future. For this reason, the aim of this research study is to determine the papers published in six major academic journals in the field of education sciences in Turkey and to synthesize the present situation by examining them in terms of certain criteria. With this aim, this study was designed to answer the following research questions:

In accordance with main themes identified in the field of science (biology, chemistry and physics), science-technology and mathematics education;

1) In which fields of education were the scale development studies conducted mostly?
2) What are the numbers of citations that the scale development studies recieved?
3) What kinf of samples were frequently used in the scale develpment studies?
4) Which sample sizes were frequently used in the scale develpment studies?
5) What kinds of validity analysis were frequently used in the scale develpment studies?
6) Which stages of construct validity were frequently used in the scale develpment studies?
7) What kinds of reliability analyses were frequently used in the scale develpment studies?

Method

This research is a study of thematic content analysis (meta-synthesis). Thematic content analysis studies are the studies which examine, synthesize and interpret the results of more than one research conducted on same topic within the framework of created themes or templates with a critical perspective unlike the raw data (Au, 2007; Finfgeld, 2003; Walsh & Downe, 2005). Thanks to this property; in that, because of synthesizing and exemplifying the similarities of the studies which examine the different dimensions of a specific subject, thematic content analysis also becomes a valuable reference resource for researchers, teachers and policy makers in terms of accessing to more studies (Çalık, Ayas & Ebenezer, 2005; Çalık & Sözbilir, 2014). Briefly, thematic content analysis include dealing with the studies done in a specific field with a qualitative approach and presenting the similarities and differences contrastively. Therefore, the number of studies (sample size) included in a meta-synthesis is generally limited as compared with those of meta-analysis and descriptive content analysis (Çalık & Sözbilir, 2014).

The Scope and Process of Research

Data Collection Process

In this research, it was previously decided to review the papers published in six major academic journals in Turkey during the years 2000-2013. As selection criteria, the journals that were national, indexed in Social Science Citation Index [SSCI] of Thomson Reuters® (Eurasian Journal of Educational Research, Education and Science, Hacettepe University Journal of Education, Educational Sciences: Theory & Practice), focused on sciences (International Journal of Environmental and Science Education and Journal of Turkish Science Education) were selected. In addition, these journals were indexed in ERIC database. As stated by Acar-Güvendir and Özer-Özkan (2015), the main reason for this is that the papers in these databases are exposed to an assessment process in accordance with criteria determined by international standards and thus more reliable results may be obtained when the papers in these journals are taken into account.

The papers which were free and open access were downloaded from web pages of journals and the other ones that were not free obtained from the library of Atatürk University. As a result, the full texts of over 200 papers about scale development/adaptation were collected. These papers were re-reviewed in accordance with criteria of inclusion and exclusion from the research and thus a total of 22 articles were included in the research. The codes and detailed information about papers are listed in Appendix 1.

The Criteria of Inclusion and Exclusion from the Research

- **Publication in selected journals:** It was taken into account that the papers were published in journals selected according to pre-determined criteria.
- **Publication in limited years:** It was taken into account that the papers were published during the years 2000-2013.
- **Publication about scale development:** In this research of thematic content analysis, only scale development studies were reviewed. Therefore, the studies of test development or scale adaptation were not included in this research. In additon, the studies of questionnaire development but not scale development were not included in the analysis.
- **Suitability to the research area:** The scale development studies in the field of biology, physics, chemistry, science-technology and mathematics education were included in the analysis.
- **Suitability of sample:** It was taken into account that the sample included in thematic content analysis was selected only from Turkey.

The Topics Consisting of Main Themes of the Studies Included in Thematic Content Analysis and Method of Coding

In the research, the authors worked together for coding of papers that were to undergo thematic content analysis process. Thus, main themes and sub-themes were identified by examining the papers one by one. And then, identified main themes and sub-themes that were reviewed with all aspects in detail were coded and presented in tables (Table 1, Table 2 and Table 3).

Table 1. The Codes and Main Themes of Papers

Codes of Main Themes	**Themes**
AISDS	Attitude, Interest Scale Development Studies
SESDS	Self-efficacy Scale Development Studies
ASDS	Anxiety Scale Development Studies
OSDS	Other Scale Development Studies

After determining main and sub-themes, in order to provide the reliability of research, both researchers selected a paper from each of the determined main themes randomly and examined them individually. Based on the self-identification of the researchers, the data were compared and researchers were found to be unanimous for majority of examined papers. In case of any inconsistency, the papers were reviewed by researchers together and the inconsistencies were resolved.

In addition to reliability, the validity also tried to be ensured in this study. For this aim, the research was implemented in the framework of three types of validity for meta-synthesis defined by Sandelowski and Barroso (2007) as following (cited in: Aküzüm & Özmen, 2013):

1. Descriptive validity: It is a type of validity which identifies the accuracy of the data gorunded on the facts. These are the meaningful and accurate identifications obtained from each of the reports used in the study.
2. Interpretive validity: It provides representing the perspectives of the researchers related to point of views completely and accurately.
3. Theoretical validity: It refers to the researchers' reliability in the interpretation of findings. This means depending on the method used to interpret the data in information coupling.

Results

In this study, the papers included in thematic content analysis were ranked according to the fields and publication years under the created main themes. Thus, it was coded the biology education papers as "B1, B2…", the physics education papers as "P1, P2…", the chemistry education papers as "C1, C2…", the science-technology education papers as "ST, ST2…" and the mathematics education papers as "M1, M2…". Therefore, these code ranges were taken into consideration in the analysis. And then, it was respectively presented in tables the findings from field, publication year, number of citation, type of sample, sample size, validity and reliability analysis for each of main themes (Table 2). In addition, the classification system presented in Worthington and Whittaker (2006) for determining the stages of construct validity were utilized (Table 3). The findings from analysis were presented as follows:

As can be seen in Table 2, there are 13 papers related to the main theme of AISDS and it is understood that B1 and B2 coded papers received the most of citations. It is also understood that other papers received very little or no citations (ST4). Regarding SESDS, there were 3 papers and M3 coded paper from this main theme received many citations, but M5 coded paper did not receive any citation. Similarly, the findings from the main theme of ASDS indicate that there were only 2 papers and one of these papers received 3 citations but the other one did not receive any citation. Regarding OSDS, it was determined there were 4 papers related to this main theme and one of these papers (M7) received more citations, but the other one received very little or no citations (M9).

The findings in terms of the sample group displays that the studies were mostly implemented with secondary school and undergraduate students and these were followed by educators and primary school students. Besides, the findings in terms of sample size indicated that 301-500 samples were preferred in approximately half of the papers and the samples under the 300 persons were also preferred in half of the rest of papers.

Regarding the validity analysis, the construct validity was examined in all of the papers (Table 2) and factor analysis was used (Table 3). Moreover, it is clearly seen that 8 papers were included in discriminant validity and more than half of papers were also used content validity. However, despite the fact that the criterion validity was included in only one paper, it is no clearly understood whether more than the half of the papers were used face validity or not. Regarding the findings from reliability analysis, it is seen that all of the papers used the internal consistency and also, approximately one-third of the papers preferred split half method. It is also noteworthy that only 2 papers included test-retest method and parallel-forms method was not used.

When the findings related to construct validity was investigated in detail, it became evident that all of the papers used the exploratory factor analysis (EFA) and few papers used confirmatory factor analysis (CFA) of the types of factor analysis (Table 3). Regarding the findings from EFA, Barlett's test of sphericity and Kaiser-Meyer-Olkin test of sample adequacy were mostly preferred for criteria used to assess factorability of correlation matrix and orthogonal rotation was the most commonly preferred rotation method. On the other hand, the findings related to criteria for item deletion/retention indicated that only one paper was included in factor loadings and more than half of the papers were taken into account cross-loadings and also, most of papers frequently preferred item analysis. In addition, it is understood that only 4 papers were included in communalities. Regarding the criteria for factor deletion/retention, it was found that despite the fact that more than half of the papers took eigenvalues, scree plot and minimum proportion of variance accounted for by factors into consideration, number of items per factor was included in a very limited number of papers.

Table 2. Analysis of Studies in Terms of Main Themes

Main themes	Code of paper	Publication year	Citations	Sample					Sample size						Validity analysis				Reliability analysis			
				Primary (1-8)	Secondary (9-12)	Undergrad.	Educator	Other	0-150	151-300	301-500	501-750	751-1000	Over 1000	Content val.	Criterion val.	Construct val.	Face val.	Test re-test	Parallel forms	Internal consistency	Split half
AISDS	B1	2002	41	-	-	-	✓	-	✓	-	-	-	-	-	?.	-	✓	?.	-	-	✓	-
	B2	2006	58	-	✓	-	-	-	-	-	-	-	✓	-	?	-	✓	?	-	-	✓	✓
	B3	2007	7	-	✓	-	-	-	-	✓	-	-	-	-	✓	-	✓	✓	-	-	✓	✓
	B4	2008	4	-	-	✓	-	-	-		✓	-	-	-	?	-	✓	?	-	-	✓	-
	F1	2010	3	-	✓	-	-	-	-	✓	-	-	-	-	✓	-	✓	?	-	-	✓	-
	F2	2012	1	-	✓	-	-	-	-	✓	-	-	-	-	✓	-	✓	✓	-	-	✓	-
	K1	2012	2	-	✓	-	-	-	-	-	✓	-	-	-	✓	-	✓	?	✓	-	✓	-
	FT1	2006	4	-	-	✓	-	-	-	-	✓	-	-	-	✓	-	✓	✓	-	-	✓	-
	FT2	2009	3	✓	-	-	-	-	-	-		-	✓	-	✓	-	✓	?	-	-	✓	-
	FT3	2009	9	-	-	✓	-	-	-	-	✓	-	-	-	✓	-	✓	?	-	-	✓	✓
	FT4	2013	-	✓	-	-	✓	✓	-	-	-	-	-	✓	✓	-	✓	-	-	-	✓	✓
	M1	2006	1	-	-	✓	-	-	-	✓	-	-	-	-	?	-	✓	?	-	-	✓	✓
	M2	2008	9	-	-	✓	-	-	-	-	✓	-	-	-	✓	-	✓	?	-	-	✓	-
	Total		**142**	**2**	**5**	**5**	**2**	**1**	**1**	**4**	**5**	**0**	**2**	**1**	**9**	**0**	**13**	**3**	**1**	**0**	**13**	**5**
SESDS	M3	2003	33	✓	-	-	-	-	✓	-	-	-	-	-	-	-	✓	-	-	-	✓	-
	M4	2007	16	✓	-	-	-	-	-	-	-	✓	-	-	✓	-	✓	-	-	-	✓	✓
	M5	2010	-	-	✓	-	-	-	-	-	✓	-	-	-	-	-	✓	-	-	-	✓	-
	Total		**49**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**1**	**1**	**0**	**0**	**1**	**0**	**3**	**0**	**0**	**0**	**3**	**1**
ASDS	FT5	2013	3	✓	-	-	-	-	-	-	-	✓	-	-	✓	-	✓	?	✓	-	✓	✓
	M6	2008	-	-	-	✓	✓	-	-	-	-	-	-	✓	✓	✓	✓	?	-	-	✓	-
	Total		**3**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**2**	**1**	**2**	**0**	**1**	**0**	**2**	**1**
OSDS	K2	2012	1	-	-	-	✓	-	-	-	✓	-	-	-	?	-	✓	✓	-	-	✓	-
	M7	2002	6	-	-	✓	-	-	✓	-	-	-	-	-	?	-	✓	?	-	-	✓	-
	M8	2013	2	-	✓	-	-	-	-	-	-	✓	-	-	✓	-	✓	✓	-	-	✓	-
	M9	2013	-	-	-	✓	-	-	-	-	✓	-	-	-	?	-	✓	?	-	-	✓	-
	Total		**9**	**0**	**1**	**2**	**1**	**0**	**1**	**0**	**2**	**1**	**0**	**0**	**1**	**0**	**4**	**2**	**0**	**0**	**4**	**0**
Overal Total			**203**	**5**	**7**	**8**	**4**	**1**	**3**	**4**	**8**	**3**	**2**	**2**	**13**	**1**	**22**	**5**	**2**	**0**	**22**	**7**

Table 3. Stages of construct validity in studies

Main themes	Code of paper	EFA	CFA	Criteria used to assess factorability of correlation matrix				Rotation methods				Criteria for item deletion/ retention				Criteria for factor deletion/ retention			
				Participants per item	Barlett's test of sphericity	Kaiser-Meyer-Olkin	Other	Unspecified/ no used	Orthogonal	Oblique	Unspecified/ no used	Loadings	Cross-loadings	Communalities	Item analysis	Eigenvalues	Scree plot	Min. variance explained by the factor	Number of items per factor
AISDS	B1	✓	-	-	✓	✓	-	-	✓	-	-	✓	-	-	-	✓	✓	✓	-
	B2	✓	-	-	✓	✓	-	-	✓	-	-	✓	-	-	✓	✓	✓	✓	-
	B3	✓	✓	-	✓	✓	-	-	-	✓	-	✓	-	-	✓	✓	✓	✓	✓
	B4	✓	-	-	✓	✓	-	-	✓	-	-	✓	✓	-	✓	✓	✓	✓	-
	P1	✓	-	-	✓	✓	-	-	✓	-	-	✓	✓	✓	-	✓	✓	-	-
	P2	✓	-	✓	✓	✓	✓	-	✓	-	-	✓	✓	-	-	✓	✓	✓	✓
	C1	✓	-	-	✓	✓	-	-	✓	-	-	✓	✓	-	✓	✓	✓	-	-
	ST1	✓	-	-	✓	✓	-	-	✓	-	-	✓	-	-	✓	✓	✓	✓	-
	ST2	✓	-	-	✓	✓	-	-	✓	-	-	✓	✓	-	-	-	✓	✓	-
	ST3	✓	✓	-	✓	✓	-	-	✓	-	-	✓	✓	-	✓	✓	✓	✓	-
	ST4	✓	-	✓	✓	✓	✓	-	✓	-	-	✓	-	-	✓	✓	✓	✓	-
	M1	✓	-	-	-	-	-	✓	✓	-	-	✓	-	-	✓	✓	-	-	-
	M2	✓	-	-	✓	✓	-	-	-	-	✓	✓	✓	-	✓	-	✓	✓	-
	Total	**13**	**2**	**2**	**12**	**12**	**2**	**1**	**11**	**1**	**1**	**13**	**6**	**1**	**9**	**11**	**12**	**10**	**2**
SESDS	M3	✓	-	-	-	-	-	✓	-	-	✓	-	-	-	✓	-	-	-	-
	M4	✓	-	-	✓	✓	✓	-	✓	-	-	✓	✓	-	-	✓	✓	-	-
	M5	✓	✓	-	-	-	-	✓	-	✓	-	✓	✓	-	✓	-	-	✓	-
	Total	**3**	**1**	**0**	**1**	**0**	**1**	**2**	**1**	**1**	**1**	**2**	**2**	**0**	**2**	**1**	**1**	**1**	**0**
ASDS	ST5	✓	-	-	✓	✓	✓	-	✓	-	-	✓	-	-	✓	✓	✓	✓	-
	M6	✓	-	-	✓	✓	✓	-	✓	-	-	✓	-	-	✓	✓	-	-	-
	Total	**2**	**0**	**0**	**2**	**2**	**2**	**0**	**2**	**0**	**0**	**2**	**0**	**0**	**2**	**2**	**1**	**1**	**0**
OSDS	C2	✓	✓	-	✓	✓	-	-	✓	-	-	✓	✓	-	✓	✓	✓	-	✓
	M7	✓	-	-	✓	✓	-	-	-	-	✓	✓	✓	✓	✓	✓	✓	✓	-
	M8	✓	-	✓	✓	✓	-	-	✓	-	-	✓	✓	✓	✓	-	✓	✓	-
	M9	✓	✓	✓	✓	✓	-	-	-	✓	-	✓	✓	✓	✓	✓	✓	-	-
	Total	**4**	**2**	**2**	**4**	**4**	**0**	**0**	**2**	**1**	**1**	**4**	**4**	**3**	**4**	**3**	**4**	**2**	**1**
Overal Total		**22**	**5**	**4**	**19**	**18**	**5**	**3**	**16**	**3**	**3**	**21**	**12**	**4**	**17**	**17**	**18**	**14**	**3**

Discussion, Conclusion and Suggestions

This research revelaed that there were many studies dealing with scale development/adaptation in Turkey, but the majority of scale development studies in the field of biology, physics, chemistry, science-technology and mathematics education were performed mostly in the field of mathematics education. Despite the fact that it is difficult to do an acceptable explanation about the reason for which the scale development studies in the field of mathematics education were more than those in the field of the science education, this result may be due to the fact that the analyses related to scale development require advanced statistical knowledge and mathematics educators have basic knowledge and skills required for statistical process due to their fields.

When reviewed papers were evaluated as to main themes, it can be seen that more than half of the papers focused on developing attitude scales and there was a limited number of papers on self-efficacy, anxiety and other subjects. There is a similar situation for the number of citations. A study by Tosun ve Taşkesenligil (2014), which was done as a document analysis of the developed/adapted scales in the field of science education in Turkey, got similar findings. The attitudes, which are important and critical predictors of individual behaviors, can be considered as more comprehensive than other psychological constructs such as anxiety, self-efficacy, and so on due to its inclusion of cognitive, affective, and behavioral dimensions (Anderson, 1988). Content analysis studies conducted in science and mathematics education (Çiltaş vd., 2012; Sozbilir, Kutu ve Yasar, 2012) indicate that the attitudes were among the most widely researched variables except for the learning, teaching basic subject areas. And also, its being a widely examined variable requires the attitude scale to be developed in the scope of many different courses. Therefore, this situation may cause that scale development studies were more focused on attitudes.

The findings related to citations in terms of AISDS indicated that B1 and B2 coded papers received the most of citations. Considering the teachers'attitudes towards laboratories, it is a natural result that B1 coded paper was used in a lot of studies which try to determine the attitudes towards teachers and laboratory courses that were among the essential parts of science teaching process. Besides, the fact that publication year is a very old time may cause that B1 coded paper received the number of citations. There is also a similar situation for M3 coded paper in main theme of SESDS and M7 coded paper in main theme of OSDS. The fact that B2 coded paper received a lot of citations may be caused by being on environment. In fact, the studies by Gul and Sozbilir (2015) indicated that environment education was the most popular research subject studied in the field of biology education in both Turkey and abroad. Gul and Sozbilir (2015) stated that this result was most likely due to the fact that environment subject is an interdisciplinary topic studied by different researchers not only by biology educators. Regarding ASDS, it is remarkable that ST5 coded paper received more citations than other papers under the main theme of ASDS although it was published in a newer date. This result may be caused that this paper addresses to a large sample group due to its inclusion the scales towards teachers, students and parents and thus it was used in more studies.

When the findings in terms of the sample group were examined, it was found that the papers included in thematic content analysis were mostly implemented with secondary school and undergraduate students. Similar findings by Tosun and Taşkesenligil (2014) may be caused due to researchers' preference of samples being reached easily. When it was considered that papers were conducted with the prospective teachers, many of whom were undergraduate students, these findings can be interpreted in the way that the researchers preferred this sample group due to its reaching

convenience. These findings and interpretations are consistent with the findings of studies on content analysis of research papers (Çiltaş et al., 2012; Gül & Sözbilir, 2014). On the other hand, limited number of sample diversity may imply that the basic components of the teaching process (students, teachers, parents, administrators etc.) were generally ignored. In addition, when it is considered that properties such as attitudes, self-efficacy, anxiety etc. can be influenced by not only oneself but also other surrounding elements and people (Deveci, Çalmaz & Açık, 2012; Gençtürk & Memiş, 2010), that the sample used in the future studies is to be spread over a broad base may contribute to obtain more reliable and accurate results for determining and solving the present problems.

On the other hand, the findings in terms of sample size indicated that 301-500 samples were preferred in approximately half of the papers and the samples under the 300 persons were also preferred in half of the rest of papers. It is a positive situation that the number of samples is usually kept as 301-500 in terms of validity and reliability but it seems necessary to increase the sample size for obtaining more reliable results. Although Comrey and Lee (1992) stated the number needed for sample size as 100= poor, 200= medium, 300= best, 500= very good and 1000= excellent, Alemo (1976) stated that the lower limit should be 400. Moreover, the fact that criteria for participants per factor in EFA were in a low level and also the findings by Tosun and Taşkesenligil (2014) indicated that sample size generally ranged between 101-200 may imply that there is a reliability problem in terms of sample size in the scale development studies in Turkey. However, Delice (2010) stated similar opinions by emphasizing that sample size is an important factor for reliability analysis as in case that a very small sample size is selected, power of test can decrease. On the other hand, it is not always a correct way to think that increasing the sample size is a way solving the reliability problems as to what extent is the sample is appropriate and how accurately the forms given to the participants have been completed are also important issues that should not be underestimated. However, any explanation about this situation related to reliability was not found in scale development studies.

When the findings from validity analysis of developed scales are examined, it seems that only one paper used the criterion validity. As the criterion validity, due to its nature, is mostly preferred in the process of test development in the field of education, this finding is a usual situation in terms of this research. However, the results of research indicate that some problems about the testing of content validity in the papers published since 2000 continue still and are losing up to date. Similarly, a research by Slavec and Drnovsek (2012) indicated that very few studies that used the content analysis used the content validity. In this research, while approximately half of the papers included the analysis and findings related to content validity, there was no clear information related to the content validity in the majority of remaining papers. There is also similar result for the findings of face validity. The reason for which- there was no clear information related to the face and content validity- may be that in some studies the expert opinion was directly taken for content validity in the stage of scale items' preparation, but in some other studies the expert opinion was taken for the purposes such as language, clarity of expressions, measuring the behavior or not etc. As stated by Brinkmann (2009), like face validity, content validity is a consensus issue and thus, for content validity, experts have to agree that the construct has been operationalised capturing all facets of the construct. According to this, it is thought that the researchers directly applied to the expert opinion for both types of validity. In parallel, Tavşancıl (2002) stated that face validity was generally assessed under the content validity and validity level in this type of validity is determined via expert opinion but not via numerical values. On the other hand, content validity, which is described as the degree of service to purpose of the test as a whole and each of items in test (Tekin, 1982, p: 45), is usually one of the basic steps which

need to be taken in the early stages of study. However, face validity, which is described in the way that what it superficially appears to measure (Ercan & Kan, 2004), is tested after the developmen of the scale. Therefore, the fact that some papers dealt with the face validity of scale in the early stages and some papers described processes of the face validity as content validity may imply that the researchers were exactly able to describe and even confuse the both types of validity. Moreover, it seems that none of the papers included in thematic content analysis were applied to statistical techniques such as content validity ratio or content validity index and, the content validity was only examined with logical paths by applying expert opinion. Therefore, these findings disclosure that researchers should develop themselves more about validity analysis.

On the other hand, regarding validity analysis, it seems that the construct validity was examined in all of the papers included in the thematic content analysis. In addition, although a lot of methods are suggested for providing the construct validity, most frequently used methods are factor analysis, discriminant validity and convergent validity (Çokluk, Şekercioğlu & Büyüköztürk, 2014). As for the findings related to the construct validity, it was found that only 8 papers used discriminant validity and convergent validity wasn't used. To Churchill (1979), the construct validity in a measurement tool requires that there is discriminant validity which displays a low correlation between variables measuring that construct and variables measuring other constructs. In this perspective, it seems that the researchers ignored the discriminant validity in the scale development studies. The reason for which- the researchers didn't frequently prefer these two types of validity- may be that they had no sufficient knowledge about these types of validity and factor analysis is more popular as construct validity and researchers may consider this analysis as sufficient. Thus, as examining the findings from this research, it is understood that factor analysis, which is one of the most frequently used methods for investigating construct validity, was used in all of the papers. Erkuş (2012) emphasizes that factor analysis is a statistical process which should absolutely be applied in the process of psychological scale development. In this perspective, it is a satisfactory situation in terms of research that factor analysis was preferred in reviewed papers. Regarding factor analysis, it seems that exploratory factor analysis (EFA) was used in all of the papers and confirmatory factor analysis (CFA) was used only in 5 papers. Similar findings were found in the studies of Hinkin (1995), Tosun and Taşkesenligil (2014). As known, EFA aims at realizing present psychological structure and CFA aims at testing this structure. Hence, both of these analyses are important in the process of scale development and they complement each other (Erkuş, 2012). One of the weaknesses of the typical factor analytic techniques is the resulting factor structure is the inability to demonstrate the goodness of fit (Long, 1983). Therefore, the analysis should be started with EFA to assess the underlying factor structure and then analysis should be followed by CFA using a different sample (or samples) to evaluate the EFA-informed structure (Cabrera-Nguyen, 2010; Worthington & Whittaker, 2006). For this reason, the fact that majority of the studies didn't include CFA may stem from researchers' lack of knowledge and skills on this subject. Besides, although EFA can be conducted with statistical programs such as SPSS which are commonly used by researchers, less well-known programs such as LISREL and AMOS need to be conducted for CFA. A study by Hinkin (1995) indicated that usage of LISREL was preferred less than SPSS in the analysis of scale development. In parallel, it can be said that the researchers mostly ignored CFA due to their inability to use these programs.

When the findings from EFA are examined in detail, regarding factorability of correlation matrix, there was no clear information in only 3 papers. In addition, majority of the papers used Barlett's test of sphericity and Kaiser-Meyer-Olkin test of sample adequacy together. In addition to

these tests, in only 4 papers participants per item were taken into account. Similar results were found in a study by Worthington ve Whittaker (2006). Moreover, Erkuş (2012) stated that factor analysis was actually based on correlation between the items and because of the fact that the correlation was very sensitive to the number of observations in the sample, factor analysis was influenced with the size of the sample. This situation implies that sample should be also taken into account for determining the factorability of correlation matrix. However, it is noteworthy that participants per item were taken into acoount in few papers. This situation may imply that the researchers consider sufficient K.M.O test, which was used in adequacy test of sampling, for this stage of research. In addition, the fact that all criteria of factorability are taken into account in the studies of scale development may be useful in behalf of reaching accurate results for next stages of scale development process.

When the rotation methods used in the stages of construct validity is examined, it is seen that majority of the papers made rotation process. As rotation process enables the factors to be interpreted more easily, it is natural that this method was used in majority of the studies. On the other hand, two main types of rotation are used: namely orthogonal and oblique. The findings from this research indicated orthogonal rotation (especially varimax) was used in majority of papers. Büyüköztürk (2002) stated both types of rotation produce similar results but orthogonal rotation is more preferred because of the fact that it provides convenience of interpretation in nearly all applications. In this perspective, it is an expected situation that orthogonal rotation was frequently used in reviewed papers.

Regarding criteria for item deletion/retention in EFA, Worthington ve Whittaker (2006) stated that item deletion is a very common and expected part of the process and also emphasized that researchers most often use the values of the factor loadings and cross-loadings on the factors to determine whether items should be deleted or retained. Similar to Worthington ve Whittaker (2006), this research founde that majority of the papers used factor loadings and cross-loadings. However, the limited number of studies included in communalities. As factor loadings indicate coorelations of items with related factors, these values are taken into account for determining under which factor items are included in scale development process (Erkuş, 2012, s:98). Therefore, it can be thought as a positive situation that researchers did not ignore these two properties in their studies. Although it may be relatively reduced to 0.30 in practice for few items, Büyüköztürk (2002) suggests that lower limit is 0.45 or higher is a measure for selection. Besides, deleting items before establishing the final number of factors could actually reduce the number of factors retained. On the other hand, unnecessarily retaining items that fail to contribute meaningfully to any of the potential factor solutions will make it more difficult to make a final decision about the number of factors to retain (Worthington ve Whittaker, 2006). Therefore, in subsequent studies similar to this one, it is thought to be useful to examine the lower limit determined by researchers during deletion of items. In addition to those mentioned above, it seems that item analysis was used in more than half of the papers in terms of criteria for item deletion/retention. The reason that item analysis is used in Likert scale is to provide one-dimension structure which is the most important assumption of Likert scaling techniques (Tavşancıl, 2002). However, Erkuş (2012) emphasized that today it was not significant to apply to factor analysis techniques for only construct validity and item analysis should absolutely be conducted in factor analysis. Therefore, it is suggested that item analysis should absolutely be included in the future studies dealing with scale development.

In this research, criteria for item deletion/retention were also examined regarding EFA. The findings indicated that more than half of the papers used eigenvalues and scree plot. Hinkin (1998) stated that eigenvalues of greater than 1 and a scree test of the percentage of variance explained

should be used to support the theoretical distinctions in the studies of construct validity and supported the findings of this research. Besides, the findings from this research indicated that more than half of the papers took minimum proportion of variance accounted for by factors into account. Çokluk et al (2014) suggest that reserachers should not make a decision on item deletion by referring to the results of the common factor variance and such a problem should be observed in other analysis (factor loading, eigenvalue etc.). According to this, considering the other properties mentioned above in terms of factor deletion disclosure that the researchers did an accurate preference. On the other hand, the number of items per factor was generally ignored in terms of criteria for factor deletion/retention. This finding may result from that the researchers did not need to provide required information in this subject due to the fact that number of items per factor were sufficient. Similarly, Erkuş (2012) stated that for determining the number of factor, not only to stay connected with the relative difference of eigenvalues and scree plot but also to analysis by taking into account conceptual structure such as all possible kinds of factor analytical techniques, the relation of arising infrastructures with together etc. Therefore, it seems important and required to conduct the analyses by considering all possible criteria in the stage of factor deletion/retention in a scale development study.

In addition to those mentioned above in this research, the reliability analysis methods in the papers included in thematic content analysis was also examined. The findings indicated that all of the papers used the internal consistency (C. Alpha) and approximately one-third of the papers preferred split half method. The findings also indicated that few papers were included in test-retest method and parallel-forms method was not used. As known, internal consistency and split half methods are reliability methods based on one application but test-retest method and parallel-forms are reliability methods based on two applications and they require twice application of the same form. As can be seen, methods based on two applications require more time, labor, expenditure etc. This situation may cause that researchers frequently preferred reliability methods based on one application. In addition, usage of internal consistency method in all papers for reliability analysis can be considered as a right choice. In fact, Tezbaşaran (1996) similarly stated that one of the basic assumptions related to the structures of Likert type scale is that each of items in scale is in a monotonic relationship with measured property that is, each of items measures same property. Therefore, firstly the internal consistency (Cronbach α) should be tested in Likert type scales.

Based on the results stated above, the following suggestions may be recommended:

- The findings obtained from thematic content analysis indicated that the studies were mostly focused on developing attitude scale. Therefore, in the framework of the needs identified, more studies are needed in subjects such as self-efficacy, anxiety, perception and so on.
- The findings indicated that the scale development studies were mostly conducted in the field of mathematics education. Therefore, biology, physics and chemistry education researchers should be directed to do more scale development studies in their field themselves.
- The scope of scale development studies should be expanded by including different samples rather than addressing a specific target group.
- The findings from research indicated that studies included a moderate sample size in general. Therefore, it should be worked with larger samples to develop more reliable scales.
- It should be provided the researchers to obtain more detailed information about different validity and reliability methods.

- Graduate courses or in-service training should be given to researchers to be able to do CFA analysis about usage of different statistical programs such as LISREL in addition to SPSS.
- It should be taken into account all criteria for factor deletion/retention and according to this, the analysis should be conducted by evaluating the findings as a whole.
- Because of the fact that CFA in reviewed studies was in very limited number, detailed examination of CFA was not carried out. Therefore, it is suggested that more detatiled examinations related to CFA should be done in the future studies.
- Finally, taking into account the criteria of inclusion and exclusion from the research, some situations, for instance, the fact that this study was conducted in the field of only science, science and technology and mathematics education; especially the papers and thesis in YOK and ULAKBIM databases were excluded from the research due to examining the journals indexed in specific databases; only scale development studies by excluding scale adaptation studies were examined, are thought as limitations of this study. Therefore, it is suggested that these limitations should be taken into account in the future studies and thus the scope of study should be more expanded.

References

Acar-Güvendir, M., & Özer-Özkan, Y. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi [The examination of scale development and scale adaptation articles published in Turkish academic journals on educatıon]. *Elektronik Sosyal Bilimler Dergisi, 14*(52), 23-33.

Akgöz, S., Ercan, İ., & Kan, İ. (2004). Meta-analizi [Meta-analysis]. *Uludağ Üniversitesi Tıp Fakültesi Dergisi, 30*(2), 107-112.

Aküzüm, C., & Özmen, F. (2013). Eğitim denetmenlerinin rollerini gerçekleştirme yeterlikleri bir meta-sentez çalışması [The efficacies of edicationel supervisors in performing their supervisory roles: A meta-synthesis study]. *EKEV Akademi Dergisi, 17*(56), 97-120.

Aleamoni, L. M. (1976). The relation of sample size to the number of variables in using factor analysis techniques. *Educational and Psychological Measurement, 36*, 879-883.

Anderson, L. W. (1988). Attitudes and their measurement. In J. P. Keeves, (Ed.), *Educational research, methodology and measurement: An international handbook* (pp.421-426). New York, Pergamon Press.

Azaltun, M. (2008). VI. araştırma yöntemleri semineri - Ölçme ve ölçek geliştirme [VI. research methods seminar - Measurement and scale development]. *Anatolia: Turizm Araştırmaları Dergisi, 19*(1), 104-111.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*, 258-267.

Brinkman, W. P. (2009). Design of a questionnaire instrument. In S. Love (Edt). *Handbook of mobile technology research methods* (pp.31-53). London: Nova Publisher.

Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı [Factor analysis: basic concepts and using to development scale]. *Kuram ve Uygulamada Eğitim Yönetimi, 32*, 470-483.

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.

Churchill, G. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*(1), 64-73.

Comrey, A. L., & Lee, H. L. (1992). *A first course in factor analysis*, Hillsdale, New Jersey: Erlbaum.

Çalık, M., Ayas, A., & Ebenezer, J. V. (2005). A review of solution chemistry studies: Insights into students' conceptions. *Journal of Science Education and Technology, 14*(1), 29-50.

Çalık, M., & Sözbilir, M. (2014). İçerik analizinin parametreleri [Parameters of content analysis]. *Eğitim ve Bilim, 39*(174), 33-38.

Çiltas, A., Güler, G., & Sözbilir, M. (2012). Türkiye'de matematik eğitimi araştırmaları: İçerik analizi çalışması [Mathematics education research in Turkey: A content analysis study]. *Kuram ve Uygulamada Eğitim Bilimleri,* 12(1), 515-580.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2014). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları (3. Baskı) [Multivariate statistics for the social sciences SPSS and LISREL applications (3th edit.)].* Pegem yayıncılık: Ankara.

Çüm, S., & Koç, N. (2013). Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek gelistirme ve uyarlama çalısmalarının incelenmesi [The review of scale development and adaptation studies which have been published in psychology and education journals in Turkey]. *Eğitim Bilimleri ve Uygulama, 12*(24), 115-135.

Dawis, R. V. (1987). Scale construction. *Journal of Counseling Psychology, 34*(4), 481-489.

Delice, A. (2010). Nicel araştırmalarda örneklem sorunu [The sampling ıssues in quantitative research]. *Kuram ve Uygulamada Eğitim Bilimleri, 10*(4), 1969-2018.

Deveci, S. E., Çalmaz, A., & Açık, Y. (2012). Doğu Anadolu'da yeni açılan bir üniversitenin öğrencilerinde kaygı düzeylerinin sağlık, sosyal ve demografik faktörler ile ilişkisi [The relainship of students' anxiety levels with health, social and demographic factors in a university newly opened in Eastern Anatolia]. *Dicle Tıp Dergisi, 39*(2), 189-196.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment Research & Evaluation*, *14*(20), 1-10.

Ercan, İ., & Kan, İ. (2004). Ölçeklerde güvenirlik ve geçerlik [Reliability and validity in the scales]. *Uludağ Üniversitesi Tıp Fakültesi Dergisi, 30*(3), 211-216.

Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalısmalarında karşılaşılan sorunlar [Encauntered problems in the studies of scale development and adaptation]. *Türk Psikoloji Bülteni, 13*(40), 17-25.

Erkuş, A. (2012). *Psikolojide ölçek ve ölçek geliştirme-I: Temel kavramlar ve işlemler (1. Baskı) [Scale and scale development in psychology-I: Basic concepts and processes]*. Ankara: Pagem Yayıncılık.

Fassinger, R. (1987). Use of structural equation modeling in counseling psychology research. *Journal of Counseling Psychology, 34*(4), 425-436.

Finfgeld, D. L. (2003). Metasynthesis: The state of the art-so far. *Qualitative Health Research*, *13*(7), 893-904.

Fraenkel, J. R., & Wallen, N. E. (2000). *How to design & evaluate research in education* (4. baskı). London: McGraw Hill.

Gençtürk, A., & Memiş, A. (2010). İlköğretim okulu öğretmenlerinin öz-yeterlik algıları ve iş doyumlarının demografik faktörler açısından incelenmesi [An investigation of primary school teachers' teacher efficacy and job satisfaction in terms of demographic factors]. *İlköğretim Online, 9*(3), 1037-1054.

Gul, Ş., & Sozbilir, M. (2015). Biology education research trends in Turkey: 1997-2012. *Eurasia Journal of Mathematics, Science & Technology Education*, *11*(1), 93-109. doi*:* 10.12973/eurasia.2015.1309a

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

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires*. Organizational Research Methods, 1*(1), 104-121. doi: 10.1177/109442819800100106

Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist*, *34*(5), 684-718.

Karadağ, E. (2009). Eğitim bilimleri alanında yapılmış doktora tezlerinin tematik açıdan incelenmesi [A thematic analysis on doctoral dissertations made ın the area of education sciences]. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi, 10*(3), 75-87.

Karagöz, Y., & Ekici, S. (2004). Sosyal bilimlerde yapılan uygulamalı araştırmalarda kullanılan istatistiksel teknikler ve ölçekler [Statistical techniques, and scales which are used in practical research in social sciences]. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *5*(1), 25-43.

Long, J. S. (1983). *Confirmatory factor analysis*. Beverly Hills, CA: Sage.

Martens, M. P. (2005). The use of structural equation modeling in counseling psychology research. *The Counseling Psychologist, 33*(3), 269-298.

McMillan, J. H., & Schumacher, S. (2009). *Research in education: Evidence-based inquiry* (7th edt.). London: Pearson.

Mor-Dirlik, E. (2014). Ölçek geliştirme konulu doktora tezlerinin test ve ölçek geliştirme standartlarına uygunluğunun incelenmesi [The analysis of the doctoral dissertations themed of scale development according to the test and scale development standards]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 5*(2), 62-78.

Quintana, S. M., & Maxwell, S. E. (1999). Implications of recent developments in structural equation modeling for counseling psychology. *The Counseling Psychologist, 27*(4), 485-527.

Özdamar, K. (2003). *Modern bilimsel araştırma yöntemleri [Modern scientific research methods]*. Eskişehir: Kaan Kitapevi.

Slavec, A., & Drnovsek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review, 14*(1), 39-62.

Sozbilir, M., Kutu, H., & Yasar, M. D. (2012). Science education research in Turkey: A content analysis of selected features of papers published. In D. Jorde & J. Dillon (Eds), *Science Education Research and Practice in Europe: Retrospective and Prospective* (pp.341-374). Rotterdam: Sense Publishers.

Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi [Measurement of attitudes and data analysis with SPSS].* Ankara: Nobel yayınevi.

Tekin H. (1982). *Eğitimde ölçme ve değerlendirme* (3. Baskı) *[Measurement and evaluation in education (3th edit.)]* Ankara: Mars Matbaası.

Tezbaşaran A. A. (1996) *Likert tipi ölçek geliştirme kılavuzu [Likert-type scale development guide]*. Türk Psikologlar Derneği Yayınları, Ankara.

Tinsley, H. E. A., & Tinsley, D. J. (1987). Use of factor analysis in counseling psychology research. *Journal of Counseling Psychology, 34*(4), 414-424.

Topu, F. B., Baydaş, Ö., Turan, Z., & Göktaş, Y. (2013). Öğretim teknolojisi araştırmalarında geçerlik ve güvenirlik önlemleri [Common reliability and validity strategies in instructional technology research]. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, *42*(1), 110-126.

Tosun, C., & Taşkesenligil, Y. (2014, Eylül). *Türkiye'de fen eğitimi alanında geliştirilen/adapte edilen ölçeklerin ve başarı testlerinin doküman analizi [Document analysis of scales and achievement tests developed/adapted in the field of science education in Turkey]*. XI. Ulusal Fen bilimleri ve Matematik Eğitimi Kongresi'nde sunulan sözlü bildiri, Adana.

Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: A literature review. *Journal of Advanced Nursing*, *50*(2), 204-211.

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

Yurdugül, H. (2005). Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması [Usage of content validity indexes for content validity in the studies of scale development]. XIV. Ulusal Eğitim Bilimleri Kongresi, Pamukkale Üniversitesi Eğitim Fakültesi.

Yurdugül, H., & Bayrak, F. (2012). Ölçek geliştirme çalışmalarında kapsam geçerlik ölçüleri: Kapsam geçerlik indeksi ve Kappa istatistiğinin karşılaştırılması [Content validity measures in scale development studies: Comparison of content validity index and Kappa statics]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, Özel Sayı 2*, 264-271.

Appendics 1. List of papers included in research

Code of paper	Information about paper
B1	Ekici, G. (2002). Biyoloji öğretmenlerinin laboratuvar dersine yönelik tutum ölçeği (BÇLDYTÖ). *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 22*, 62-66
B2	Uzun, N., & Sağlam, N. (2006). Ortaöğretim öğrencileri için çevresel tutum ölçeği geliştirme ve geçerliliği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 30*, 240-250
B3	Arıcak, O. T., & Ilgaz, G. (2007). Açımlayıcı ve doğrulayıcı faktör analizi ile biyoloji dersi tutum ölçeğinin yapı geçerliliğinin incelenmesi. *Eurasian Journal of Educational Research, 28*, 1-8
B4	Darçın, E. S., & Güven, T. (2008). Development of an attitude measure oriented to biotechnology for the pre-service science teachers. *Türk Fen Eğitimi Dergisi, 5*(3), 72-81
P1	Tekbıyık, A., & Akdeniz, A. R. (2010). Ortaöğretim öğrencilerine yönelik güncel fizik tutum ölçeği: Geliştirilmesi, geçerlik ve güvenirliği. *Türk Fen Eğitimi Dergisi, 7*(4), 134-144
P2	Taşlıdere, E., & Eryılmaz, A. (2012). Basit elektrik devreleri konusuna yönelik tutum ölçeği geliştirilmesi ve öğrencilerin tutumlarının değerlendirilmesi. *Türk Fen Eğitimi Dergisi, 9*(1), 31-46
C1	Koçak, C., & Önen, A. S. (2012). Günlük yaşam kimyası tutum ölçeği geliştirme çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 43*, 318-329
C2	Feyzioğlu, B., Demirdağ, B., Akyıldız, M., & Altun, E. (2012). Kimya öğretmenlerinin laboratuvar uygulamalarına yönelik algıları ölçeği geliştirilmesi. *Türk Fen Eğitimi Dergisi, 9*(4), 44-63
ST1	Afacan, Ö., & Aydoğdu, M. (2006). The science technology society (STS) course attitude scale. *International Journal of Environmental and Science Education, 2*(1), 189-201
ST2	Yaşar, Ş., & Anagün, Ş. S. (2009). Reliability and validity studies of the science and technology course scientific attitude scale. *Türk Fen Eğitimi Dergisi, 6*(4), 43-54
ST3	Evrekli, E., İnel, D., Balım, A., G., & Kesercioğlu, T. (2009). Fen öğretmen adaylarına yönelik yapılandırmacı yaklaşım tutum ölçeği: Geçerlilik ve güvenirlik çalışması. *Türk Fen Eğitimi Dergisi, 6*(2), 134-148
ST4	Kağıtçı, B., & Kurbanoğlu, N. İ. (2013). Fen ve teknoloji dersine yönelik kaygı ölçeğinin geliştirilmesi: Güvenirlik ve geçerlik çalışması. *Türk Fen Eğitimi Dergisi, 10*(3), 95-107
ST5	Deveci, İ., & Önder, İ. (2013). Fen ve teknoloji derslerinde verilen ödevlere yönelik öğretmen, öğrenci ve veli ölçeklerini geliştirme çalışması. *Türk Fen Eğitimi Dergisi, 10*(3), 159-184
M1	Çağırgan-Gülten, D., & Derelioğlu, Y. (2006). Öğretmen adaylarının matematik öğrenmeyi öğretmeye ilişkin tutumlarını incelemeye yönelik bir ölçek geliştirme çalışması. *Eurasian Journal of Educational Research, 24*, 103-111
M2	Turanlı, N., Karakaş-Türker, N., & Keçeli, V. (2008). Matematik alan derslerine yönelik tutum ölçeği geliştirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 34*, 254-262
M3	Işıksal, M., & Aşkar, P. (2003). İlköğretim öğrencileri için matematik ve bilgisayar öz-yeterlik algısı ölçekleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 25*, 109-118
M4	Cantürk-Günhan, B., & Başer, N. (2007). Geometriye yönelik öz-yeterlik ölçeğinin geliştirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 33*, 68-76
M5	Özyürek, R. (2010). Matematik yetkinlik beklentisi bilgilendirici kaynaklar ölçeği'nin geçerlik ve güvenirlik çalışmaları. *Kuram ve Uygulamada Eğitim Bilimleri, 10*(1), 419-447
M6	Deniz, L., & Üldaş, İ. (2008). Validity and reliability stıdy of the mathematics anxiety scale involving teachers and prospective teachers. *Eurasian Journal of Educational Research, 30*, 49-62
M7	Çalıklıoğlu-Bali, G. (2002). Matematik öğretiminde dil ölçeği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 23*, 57-61
M8	Kayhan-Altay, & M., Umay, A. (2013). İlköğretim ikinci kademe öğrencilerine yönelik sayı duyusu ölçeği'nin geliştirilmesi. *Eğitim ve Bilim, 38*(167), 241-255
M9	Arslan, O., Işıksal-Bostan, M., & Şahin, E. (2013). Origaminin matematik eğitiminde kullanılmasına yönelik inanç ölçeği geliştirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 28(2)*, 44-57