

Examination of Scale Development and Adaptation Studies Published in the Field of Educational Sciences

Müge Vuralⁱ
Firat University

Ahmet Egemen Akmençeⁱⁱ
Ministry of National Education

Murat Tuncerⁱⁱⁱ
Firat University

Abstract

The aim of this study is to examine the scale development and adaptation studies in the subject area of educational sciences published in TR-index in 2023 in terms of subject area, rotation method, confirmatory factor analysis (CFA) fit indices, number of modifications, type of scale developed, number of items before and after exploratory factor analysis (EFA), number of dimensions, sample sizes and reliability methods. For this purpose, document analysis method, one of the qualitative research designs, was used. A total of 84 scale development and adaptation studies were accessed by examining the issues published in 2023 of 184 journals in the TR-index and listed in the subject area of "Education, educational research", and the final sample was determined as 61 studies by removing 23 scale studies that were not in the field of educational sciences. Of the identified studies, 48 were scale development and 13 were scale adaptation studies. As a result of the research, it was determined that most of the studies were prepared on the subject of competence/self-efficacy, Varimax rotation method was frequently used, CFA fit index values were appropriate, there were no modifications in most of the studies, most of the studies were prepared with a five-point Likert-type scale, the difference in the number of items before and after EFA was not very large, most of the scales consisted of three dimensions, the sample selection was generally 251-500 people and EFA and CFA samples were different, and Cronbach's Alpha method was frequently used as a reliability method. The results obtained were supported by the literature and various suggestions were made.

Keywords: Scale Development, Scale Adaptation, Scale, Scale Review

DOI: 10.29329/ijpe.2024.1055.1

Submitted: 28/03/2024

Accepted: 12/07/2024

Published: 01/08/2024

ⁱ **Müge Vural**, Educational Sciences, Institute of Educational Sciences, Firat University, ORCID: 0000-0003-1259-462X

ⁱⁱ **Ahmet Egemen Akmençe**, Ministry of National Education, ORCID: 0000-0001-7523-4061

ⁱⁱⁱ **Murat Tuncer**, Prof. Dr., Educational Sciences, Firat University, ORCID: 0000-0001-9136-6355

Correspondence: mtuncer@firat.edu.tr

INTRODUCTION

Measurement has an important place both in daily life and in scientific studies. Although it is a method of investigation used in all fields of science, measurement processes vary among scientific disciplines. While the variables observed in natural sciences are concrete and observable, they are abstract in the field of social sciences, and therefore indirect measurement is carried out in the field of social sciences. (Baykul, 2000; Özgüven, 2011). Indirect measurement is the measurement of a characteristic through another characteristic (Güler, 2011). Throughout history in the field of social sciences, the development of different theories has led to the emergence of various approaches in the field of measurement, but fundamentally, psychometric and impressionistic approaches stand out (Cronbach, 1960 as cited in Çüm and Koç, 2013). Cronbach (1960) stated that the psychometric approach is more objective, easier compared to other data collection methods, allows for objective scoring, and enables valid and reliable observations, hence its effects are seen in research across different fields (as cited in Çüm and Koç, 2013).

For a measurement, there must be the measured, the measurer, instruments, and rules managing the relationships between these three. The measured is defined as the characteristics of everything that exists and occupies space in the empirical world; therefore, things that are not directly or indirectly observed cannot be the subject of measurement (Erkuş, 2019). When a measurement instrument is developed or adapted, those conducting the measurement process must provide information regarding the psychometric properties of the scale if it is used for purposes other than its original development or with a different sample. Not providing this information will make the data obtained from this measurement process questionable (Şekercioğlu, 2023). Even the most complex analyses within the field of statistics rely on measures of similarity (central tendency) and difference (variation) (Erkuş, 2019).

In order to measure human characteristics, it is necessary to classify these characteristics as physical, psychological, biological, and physiological. Psychological characteristics are also divided into cognitive, affective, and behavioral aspects. Cognitive characteristics involve the processes of information processing following the reception of internal or external stimuli; affective characteristics encompass attitudes, interests, personality traits, and emotional features such as love, hate, like, and anxiety; and behavioral characteristics refer to observable behaviors exhibited by individuals. If the structure of psychological characteristics is found out, then the appropriate measurement methods can also be determined. In the field of psychology, scales are used to measure variables such as anxiety, attitudes, personality, motivation, values, and interests while collecting data. The measurement of concepts that cannot be directly observed or measured, but can be measured through observable variables, is defined as psychological measurement (Dirlik, 2013).

Scale development is defined as *the process of creating a set of stimuli that will only stimulate the relevant characteristics intended to be measured in an individual and forming response categories suitable for these stimuli* (Erkuş, 2019). In a developed scale, the process involves determining how much of the measured attribute is represented by the individual's responses, while scale development entails structuring the items that will reveal what these characteristics are. The starting point in scale development is theories and conceptual foundations; when these are flawed, it means the scale has also been developed erroneously. Therefore, the scale development process, as it influences the decision-making process about individuals, is a serious and demanding endeavor (Erkuş, 2019).

When the literature is examined, it is observed that there is a significant amount of work on the development and adaptation of psychological tests in our country. The lack of institutions to provide support for researchers in terms of measurement tools in our country causes researchers to experience difficulties in accessing valid and reliable scales. Therefore, they feel compelled to either adapt a scale developed abroad or develop a new scale. To develop or adapt a scale, it is necessary to have a detailed information of the subject matter and to be familiar with the standards related to the scale (Edenborough, 1999). Researchers who wish to adapt or develop a scale must have a thorough understanding of the characteristics and structure of the variable they intend to measure (Cohen and

Serdlik, 2010). Otherwise, the scale developed without considering the stages of scale adaptation or development may become a scale that poses risks for scientific research. As a result, the time and effort spent on research using this scale may go to waste, and furthermore, the information obtained from these scales, which will be included in the literature, can lead to information pollution and situations that negatively affect the lives of individuals in the research population (Çüm and Koç, 2013).

A researcher aiming to conduct scale development needs to follow eight steps (DeVellis, 2003). Firstly, they must determine what they want to measure and establish the theoretical framework related to the concept they want to measure. Secondly, they need to create an item pool. Then, in sequence, they should decide on the format of the measurement tool, have the items in the item pool reviewed by experts, conduct item validations, carry out a preliminary application of the scale, evaluate the items based on the application results, and create the final form of the scale (Şahin and Boztunç Öztürk, 2018). Various methods can be used when creating the item pool. If a directly measurable behavior is to be assessed, the observation method can be used; if theory is to be employed, a review of the literature can be conducted; if attitudes are to be measured, a sample close to the population is asked to write a composition (Erkuş, 2019). When creating an item pool, preparation should be done by considering all dimensions of the subject (Tezbaşaran, 2008). The type of measurement tool should be decided by taking into consideration how the results will be interpreted. Choosing the most suitable type of measurement tool from various types such as Thurstone, Guttman, Likert, etc., is important at this stage (DeVellis, 2003). During the stage of content validity, expert opinions are consulted to assess the representativeness and coverage validity of the items (Erkuş, 2003; Yurdugül, 2005). After determining item adequacy, validity and reliability are assessed by applying them to the sample group. It is crucial to determine the sample size very well at this stage. When a small sample size is used, internal consistency issues may arise (DeVellis, 2003). After the validity and reliability study, the scale becomes ready in its final format. Taking all these stages into account, it is obvious that scale development and scale adaptation are demanding and laborious tasks. Therefore, researchers need to be very meticulous.

The steps to be followed in scale adaptation studies have been established by examining studies by the World Health Organization. The researcher conducting scale adaptation must first translate the scale. For the translation to be done correctly, the translator must also be familiar with the culture in which the scale was developed. During translation, attention should be paid to conceptual equivalence, simple and clear expressions should be used, and the characteristics of the group that will respond to the scale should be taken into account. After translation, expert opinions should be sought to correct any missing or incorrect expressions in the translation, the cultural compatibility of the items should be examined, and the conceptual equivalence of the concepts should be scrutinized. Then, the scale should be translated into the target language by a different translator to determine the differences between the original scale and the translated one. This ensures the accuracy of the translation into Turkish. Subsequently, a pilot application is conducted on a similar group for the adaptation study, and the final version of the adaptation study is determined through analyses of this application (WHO, 2017). In scale adaptation studies, the process seeks answers to questions such as the compatibility of the adaptation to the new culture to be used, and how adequate the psychometric properties are, for the scale to be used in a culture other than the one in which it was developed (Deniz, 2007). Fundamentally, scale adaptation studies are a process that often requires repeating many of the steps applied in scale development studies (Şahin, 1994).

When studies examining scale adaptation and development in the literature were investigated; Çüm and Koç (2013) examined scale development and adaptation studies published in the TÜBİTAK Ulakbim database between 2005 and 2013, focusing on the fields of psychology and educational sciences. The research revealed that in 67% of the scale development studies, the steps of scale development were followed, while in 45% of the scale adaptation studies, the steps of scale adaptation were followed. Mor Dirlik (2014) examined doctoral theses on the topic of scale development between 2009 and 2014 to assess the compliance of the scales found in these theses with development standards. Out of a total of five studies, it was concluded that two of them met the standards at a high

level, one met them at a moderate level, and two met them at a low level. Acar Güvendir and Özer Özkan (2015) examined scale development and adaptation studies published in Türkiye between 2006 and 2014. During the examination, the necessary steps in the process were scrutinized, and similarities and differences were identified. As a result of the examination, it was found that Cronbach's Alpha was used in all scales, expert proficiency in both languages was generally employed in scale adaptation studies, adherence to scale adaptation guidelines was not observed, and similarities were generally rare in the articles. Gül and Sözbilir (2015) examined scale development studies conducted in the field of science and mathematics education between 2000 and 2013. As a result of the examination, it was found that the majority of the developed scales were attitude scales, content validity was examined as the validity analysis, and Exploratory Factor Analysis (EFA) was frequently conducted. Delice and Ergene (2015) examined scale development and adaptation studies related to mathematics education published in peer-reviewed journals between 2005 and 2014. As a result of the research, it was found that in most studies, the Cronbach's alpha coefficient exceeded 0.8, there was no significant relationship between sample size and number of items, in some studies, the number of participants per item was less than five, and in scale development studies, scale development steps were applied at a rate of 65%, while in scale adaptation studies, adaptation steps were applied in 53% of the cases.

Şahin and Boztunç Öztürk (2018) examined scales developed in the field of education in Turkey between 2010 and 2016. The scales were examined under the headings of the introduction section of the article, preparation of the item pool, developed form, sample of trial application, and analyses. As a result of the examined articles, it was found that the most common practice during item writing process was literature review, obtaining expert opinions was common, working with groups ranging from 300 to 499 individuals was prevalent, the number of participants per item was mostly between 5 to 9, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were commonly used together for establishing construct validity, criterion validity was generally not tested, item analysis included item-test correlation, and internal consistency reliability method was utilized. Tekin and Bolat (2018) examined scale development studies related to writing education published between 2006 and 2017. As a result of the examination, it was found that scales were developed on topics such as attitude, anxiety, tendency, self-efficacy, and effect. Scales are generally working with primary school students, these scales are prepared as Likert-type , and using Exploratory Factor Analysis (EFA) most frequently in data analysis, reliability analyses were also conducted. Olgun and Alatlı (2019) examined scale development and adaptation studies targeting adolescents published in Türkiye between 2004 and 2019. In the studies examined, it was found that for scale development; the most common practice during the item writing process was literature review, expert opinions from domain experts were frequently sought, pilot testing was conducted, the number of participants per item was mostly 20 and above, item-test correlation was used for item analysis, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were used together for validity, and Cronbach's alpha was most frequently used as the reliability method. In scale adaptation studies, it was determined that the most common practice was to consult domain experts and English language experts, pilot testing was not conducted, item-test correlation was used for item analysis, Confirmatory Factor Analysis (DFA) was mostly used as the validity method, and reliability was assessed using Cronbach's alpha and test-retest method. Çelik and Yüksel (2020) examined scales within the scope of music education developed between 1997 and 2019. As a result of the research, it was found that the most common practice was to work with professional music education institutions, predominantly single-dimensional scales were developed, the number of participants per item was insufficient, attitude scales were developed the most, and generally five points Likert-type scales were developed. Cin Şeker and Yücel Çetin (2022) examined scale development studies conducted in the field of reading education. As a result of the examination, it was found that attitude scales were developed most, the studies were conducted with middle school students, Likert-type scales were prepared, both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were used together, and reliability analysis was also conducted.

In this study, scale development and adaptation studies falling within the field of educational sciences, published in TR-index and released in 2023, were examined. The number of journals

published in TR-index is 1698. Of these journals, 184 are related to "Education, educational research". Considering the scale adaptation and scale development researches published in these journals and the hundreds of research studies conducted using these scales, it is understood that scales prepared without considering the scale adaptation and development stages would cause a significant problem. Despite this situation, it has been observed that there are few studies examining scale adaptation and development in our country. Therefore, this research aims to examine the scale adaptation or development studies published in the journals listed in TR Index under the topic of "Education, educational research" for the year 2023 according to various variables.

METHOD

Research Model

In this study, the method of document analysis, which is one of the qualitative research designs, was used as the research aimed to examine scale development and adaptation studies falling within the field of educational sciences. Detailed examinations and analyses can be conducted with qualitative research, therefore it is not suitable for working with very large sample groups (Yıldırım and Şimşek, 2013). As a result of the analyses conducted, the status of the scale development and adaptation studies was revealed according to the examined standards.

Data Collection and Data Collection Tool

The document analysis method was used to find, read, and evaluate the studies used in the research. For this study, which aims to examine scale development and adaptation studies falling within the field of educational sciences published in TR-Index indexed journals in Türkiye in 2023, journals published on the trdizin.gov.tr website were examined. The studies that will constitute the data of the research were scanned from articles falling into the field of Educational Sciences by entering the keywords "scale development" and "scale adaptation" on the websites of TR Index. A total of 84 scale development or adaptation studies were accessed, resulting in the exclusion of 23 studies that did not fall within the field of educational sciences, thus determining the final sample size as 61. Among the articles included in the analysis, 48 were scale development studies and 13 were scale adaptation studies. The examined articles were coded as S1, S2,..., S61. Although the research included 61 studies, some of them (S11, S49, and S50) contained more than one scale. Therefore, in terms of certain characteristics, the codings for these studies were given as S11A, S11B, S49A, S49B, S50A, S50B, S50C.

The examined studies were analyzed based on the purpose of the scale, EFA rotation method, CFA fit indices, number of modifications, Likert type, number of items, number of items in the item pool, difference in the number of items between the initial and final versions of the scales, number of dimensions, sample size, and reliability methods. The names of journals containing scale development and adaptation studies included in the research are provided in Table 1.

Table 1: Journals containing scale development and adaptation studies

Name of Journal	Frequency
Abant İzzet Baysal University Journal of Education Faculty	1
Adiyaman University Journal of Social Sciences Institute	1
Journal of Mother Tongue Education	3
Ankara University Journal of Faculty of Educational Sciences	2
Bartın University Journal of Education Faculty	2
Bayburt Journal of Education Faculty	2
Buca Journal of Education Faculty	4
Cumhuriyet Uluslararası Journal of Education	4
Çukurova University Journal of Education Faculty	3
Journal of Learning and Teaching in the Digital Age	2
Journal of Educational Technology Theory and Practice	1
Journal of Education and Human Sciences: Theory and Practice	1

Erzincan University Journal of Education Faculty	2
Fırat University Journal of Social Sciences	2
Gazi Journal of Education Faculty	4
Hayef: Journal of Education	2
İnönü University Journal of Education Faculty	3
Journal of Computer and Education Research	1
Kafkas Journal of Education Research	1
Kocaeli University Journal of Education	2
Marmara University Atatürk Education Faculty Journal of Educational Sciences	2
National Education	5
Muğla Sıtkı Koçman University Journal of Education Faculty	1
Pamukkale University Journal of Education Faculty	2
Sakarya University Journal of Education	2
Turkish Journal of Educational Sciences	1
Uludağ University Journal of Education Faculty	1
Van Yüzüncü Yıl University Journal of Education Faculty	1
Journal of Higher Education	1

Data Analysis

The articles used in the research were analyzed by using content analysis method. Content analysis is a method used for understanding and comparing publications. By using content analysis techniques, the contents of a study can be analyzed in depth (Arıkan, 2013). The most important stage in conducting content analysis is the creation of categories that match with the purpose of the research. Through categories, data can be transformed into meaningful information (Yıldırım & Şimşek, 2013). In this study, categorical analysis and frequency analysis were conducted. Categorical analysis and frequency analysis are often used together. Categorical analysis involves dividing content into specific parts and categorizing them according to predefined criteria (Tavşancıl & Aslan, 2001). Researchers have identified categories that align with the purpose of the study based on specific criteria, and tables have been created accordingly. Frequency distributions have been utilized to quantify the obtained data. The SPSS software package has been used for the analysis of this data.

Coding and Reporting of Data

The researchers recorded information about the articles included in the study separately in a finalized form. Subsequently, the information was compared, and discussions were held regarding any differences to reach a consensus.

Validity and Reliability

In this study, importance has been given to the following issues to ensure validity and reliability.

1. The general purpose of the study and research questions have been clearly stated.
2. The selection process of the articles constituting the data of the study has been described in detail.
3. Information about the authors, research titles, and publication details of the articles constituting the data of the study has been provided in a table.
4. To ensure the reliability of data analysis, researchers filled out separate forms and then compared the completed forms. Discussions were held regarding any discrepancies to reach a consensus.

FINDINGS

The findings regarding the distribution of topics of the examined studies are provided in Table 1.

Table 1: Distribution of topics of the examined studies

Purpose	Study Code	f
Competence/Self-efficacy	S2, S7, S25, S28, S32, S33, S43, S53, S55	9
Attitude	S17, S20, S22, S29, S35, S44, S57, S58	8
Literacy	S26, S49, S52, S60, S61	5
Assessment	S1, S12, S21, S59	4
Perception/Self-perception	S27, S39	2
Belief	S19, S51	2
Addiction	S4, S14	2
Participation	S8, S18	2
Organization/Self-regulation	S11, S15	2
Tendency	S30, S34	2
Leadership	S37, S48	2
Behaviour	S42, S45	2
Anxiety, Motivation, Strategy, Awareness, Satisfaction, Commitment, Opinion, Effective Teaching, Skill, Grade, Determination of Education Philosophy, Subject Matter Knowledge, Performans, Active Learning Practices, Immunity, Emphaty, Experiences, Approach, Metacognition	S3, S5, S6, S9, S10, S13, S16, S23, S24, S31, S36, S38, S40, S41, S46, S47, S50, S54, S56	19

When Table 1 is examined, it can be seen that the highest number of competence/self-efficacy (f=9) and attitude (f=8) scales have been developed or adapted. Single studies under sub-topics have been merged. The findings regarding rotation methods of the examined studies are provided in Table 2.

Table 2: Rotation methods of the examined studies

AFA	Study Code	f
Varimax	S2, S3, S4, S10, S12, S13, S16, S17, S18, S20, S22, S26, S29, S31, S32, S33, S36, S38, S39, S47, S49A, S52, S53, S59, S61	25
Geomin	S34, S42, S45	3
Equamax	S37	1
Oblimin	S15, S19, S23, S51, S57, S58, S60	7
Promax	S30	1
No rotation	S5, S43, S49B	3
No AFA	S1, S6, S8, S9, S11A, S11B, S14, S21, S24, S28, S40, S41, S48, S54, S56	15
No Information	S7, S25, S27, S35, S44, S50A, S50B, S50C, S55	9

When the table is examined, it is observed that studies not applying EFA are generally adaptation studies, the most used type of EFA is the Varimax rotation method (f=25), and one of the least used orthogonal rotation methods is the Equamax (f=1), and one of the oblique rotation methods is the Promax (f=1). The examined KMO values range from .79 to .98. The findings regarding the CFAs of the studies included in the research are provided in Table 3.

Table 3: CFA Information of the studies

Fit Indices	Good Fit	f	Acceptable Fit	f	Poor Fit (f)
CMIN/DF (χ^2/sd)	0-3	49	3-5	13	1
RMSEA	.00-.05	7	.05-.08	48	8
CFI	.95-1.00	40	.90-.95	24	-
GFI	.90-1.00	31	.85-.90	10	3
AGFI	.90-1.00	15	.85-.90	21	-
RMR	.00-.05	12	.05-.10	5	-
SRMR	.00-.05	21	.05-.10	22	-
IFI	.95-1.00	22	.90-.95	11	1
NFI	.95-1.00	15	.90-.95	25	4
NNFI (TLI)	.95-1.00	26	.90-.95	17	3
PGFI	.95-1.00	-	.50-.95	8	-
RFI	.95-1.00	6	.90-.95	9	1
PNFI	.95-1.00	1	.50-.95	6	-

When the table is examined, it is observed that many values (CMIN/DF, RMSEA, CFI, GFI, AGFI, SRMR, NFI, TLI) have been examined in more than half of the articles. The most frequently used values are CFI (x=64), CMIN/DF (x=63), and RMSEA (x=63), while the least used values are PGFI (x=8) and PNFI (x=7).

When classifying the fit indices, the criteria proposed by Karagöz (2017), Erkorkmaz et al. (2013) were used as a basis. The fit indices of the scales are classified as follows: for CMIN/DF, the majority (f=49) indicate "Good Fit"; for RMSEA, the majority (f=48) indicate "Acceptable Fit"; for CFI, the majority (f=40) indicate "Good Fit"; for GFI, the majority (f=31) indicate "Good Fit"; for AGFI, the majority indicate "Acceptable Fit"; for RMR, the majority (f=12) indicate "Good Fit"; for SRMR, the majority (f=22) indicate "Acceptable Fit"; for IFI, the majority (f=14) indicate "Good Fit"; for NFI, the majority (f=25) indicate "Acceptable Fit"; for NNFI, the majority (f=26) indicate "Good Fit"; for PGFI, the majority (f=8) indicate "Acceptable Fit"; for RFI, the majority (f=9) indicate "Acceptable Fit"; for PNFI, the majority (f=6) indicate "Acceptable Fit". Furthermore, it has been observed that fit indices falling outside the "Acceptable Fit" range generally have values close to the acceptable fit.

The numbers of modifications occurring in the examined studies is provided in Table 4.

Table 4: Numbers of modifications in the studies

Numbers of modifications	Study Code	f
No Information	S25, S42,	2
No Modification	S1, S2, S5, S7, S8, S9, S10, S11A, S11B, S16, S17, S19, S20, S22, S23, S24, S30, S31, S32, S33, S34, S36, S39, S43, S45, S47, S48, S49A, S50C, S51, S52, S54, S55, S58, S59	35
1-2	S6, S14, S15, S21, S28, S38, S40, S41, S44, S49B, S50A, S53	12
3-4	S3, S4, S12, S18, S26, S29, S37, S46, S56, S61	10
5-6	S50B, S57, S60	3
7-8	-	0
9+	S13, S27	2

When the table is examined, it is observed that in the majority of the studies (f=35), there were no modifications; in studies with modifications, the majority (f=12) had 1-2 modifications. Findings related to the type of scale developed in the studies are provided in Table 5.

Table 5: Type of scale developed in the studies

Likert	Study Code	f
Triple	S17, S19, S20, S49A, S49B, S51	6
Quadruple	S1, S6, S8, S11A, S18, S41	6
Quintuple	S2, S3, S4, S5, S10, S11B, S12, S15, S16, S21, S22, S23, S25, S26, S27, S28, S29, S30, S31, S32, S33, S35, S36, S37, S38, S39, S42, S43, S44, S45, S46, S47, S48, S50, S52, S53, S54, S55, S57, S58, S59, S60, S61	43
Sextuple	S24, S40	2
Septuple	S7, S13, S14, S56	4
No Information	S9	1

All of the examined scales were prepared in Likert type. When the table is examined, it is observed that scales developed or adapted as a result of the research are generally of the five-point Likert scale type (f=43). Additionally, in one article, the type of Likert scale was not specified. The distribution of scales according to the number of items is given in Table 6.

Table 6: The number of items of the scales

The number of items	Study Code	f
0-10	S7, S14, S30, S35, S44, S49B	6
11-20	S2, S4, S6, S8, S9, S12, S15, S16, S17, S19, S20, S22, S24, S29, S34, S39, S41, S42, S43, S45, S48, S49A, S51, S54, S58, S61	26
21-30	S1, S3, S5, S10, S11A, S18, S21, S23, S26, S32, S36, S37, S38, S46, S52, S53, S55, S56, S59, S60	20
31-40	S11B, S13, S25, S31, S33, S40, S47, S57	8
41-50	S27, S28, S50	3

When the table is examined, it is observed that the examined scales are divided into five categories based on the number of items. The majority of the examined scales have a number of items in the range of 11-20 (f=26). Twenty scales fall within the range of 21-30, eight scales fall within the range of 31-40, six scales fall within the range of 0-10, and three scales fall within the range of 41-50. Findings regarding the number of items created before the Exploratory Factor Analysis (EFA) in the examined scale studies are provided in Table 7.

Table 7: Number of Items before EFA

Item Pools	Study Code	f
11-20	S35, S49B	2
21-30	S7, S10, S12, S19, S29, S34, S37, S45, S51, S53	10
31-40	S3, S5, S13, S20, S38, S49A, S59	7
41-50	S16, S17, S26, S28, S30, S32, S46, S47, S61	9
51-60	S4, S23, S53, S33, S42, S55, S60	7
61-70	S2, S36, S50, S58	4
71-80	S15, S39, S57	3
81-90	S52	1
91-100	S25	1
111-120	S22	1

When the table is examined, it is observed that the examined scales are divided into 10 categories in terms of the created item pool. Additionally, it is seen that the number of created item pools is at most in the range of 21-30 (f=10). The difference in the number of items between the scales before and after the EFA is given in Table 8.

Table 8: The difference in item numbers before and after EFA.

The difference in item numbers	Study Code	F
0-10	S3, S5, S10, S13, S19, S28, S29, S35, S37, S47, S49B, S53	12
11-20	S7, S12, S31, S32, S34, S38, S45, S49A, S50, S51, S59	11
21-30	S16, S20, S23, S26, S46, S55, S61	7
31-40	S4, S17, S30, S33, S42, S60	6
41-50	S2, S36, S57	3
51-60	S39, S52, S58	3
61-70	S15, S25	2
101-110	S22	1

When the table is examined, it is observed that the examined scales are divided into eight categories based on the difference in item numbers between their initial and final versions. Additionally, it is noted that the highest frequency of item differences falls within the ranges of 0-10 (f=12) and 11-20 (f=11). The distribution of dimensions formed after EFA in the examined scales is provided in Table 9.

Table 9: The number of dimensions formed after EFA

The number of dimensions	Study Code	F
1 Dimension	S5, S7, S14, S15, S19, S25, S30, S35, S43, S49B, S55	11
2 Dimensions	S2, S4, S20, S44, S45, S48, S51, S61	8
3 Dimensions	S6, S8, S12, S16, S17, S26, S29, S32, S34, S37, S39, S41, S47, S49A, S50, S57, S58	17
4 Dimensions	S1, S9, S10, S11A, S11B, S21, S22, S23, S24, S33, S42, S52, S54, S59	14
5 Dimensions	S3, S31, S36, S38, S46, S56	6
6 Dimensions	S18, S27, S60	3
7 Dimensions	S13, S40	2
9 Dimensions	S28	1

When the table is examined, it is observed that the examined scales are divided into eight categories based on the number of dimensions. Additionally, it is noted that the maximum number of dimensions is 3. (f=17).

When the table is examined, it is observed that of all the scales, 17 of them have three-factor, 15 of them have four-factor, 11 of them have one-factor, 8 of them have two-factor, 6 of them have five-factor, 3 of them have six-factor, 2 of them have seven-factor, and 1 of them has nine-factor. The findings regarding the sample sizes of the examined scales are presented in Table 10.

Table 10: The sample sizes of the scales

Total sample sizes	EFA-CFA Sample	Study Code	f
1-250	EFA-CFA same sample	S7, S55, S59	3
	EFA-CFA different sample	-	-
	Only CFA	S8, S18, S21, S24	4
	No Information	S49A, S49B	2
251-500	EFA-CFA same sample	S20, S22, S31	3
	EFA-CFA different sample	S2, S5, S10, S12, S27, S29, S33, S38, S43, S46, S58	11
	Only CFA	S1, S6, S9, S11B, S28, S54, S56	7
	No Information	S13, S40, S41, S44, S48	5
501-750	EFA-CFA same sample	S15	1
	EFA-CFA different sample	S4, S19, S23, S25, S34, S36, S42, S45, S51, S53, S57, S60	12
	Only CFA	S11A, S14	2
	No Information	S52	1
751-1000	EFA-CFA same sample	-	-
	EFA-CFA different sample	S3, S16, S30, S32, S35, S37, S39	7
	Only CFA	-	-

1001-1250	No Information	-	-
	EFA-CFA same sample	-	-
	EFA-CFA different sample	S17, S26, S50, S61	4
	Only CFA	-	-
	No Information	S47	1

When the table is examined, it is observed that the majority of the samples consist of 251-500 individuals ($f=26$) and generally, the EFA sample differs from the CFA sample. It is noted that different samples are not used for EFA and CFA for the sample groups of 1-250 individuals. Additionally, information regarding the EFA and CFA samples was not found in nine studies. The reliability methods and distribution ranges used in the scales are provided in Table 11.

Tablo 11: The reliability methods used for the Scales

Reliability Method	Value	Study Code	f
Cronbach Alfa	$0.6 \leq \alpha < 0.7$	S44, S47	2
	$0.7 \leq \alpha < 0.9$	S1, S6, S8, S10, S11A, S11B, S14, S17, S18, S20, S22, S23, S24, S30, S32, S34, S35, S39, S41, S42, S46, S49A, S49B, S50A, S50C, S59	26
	≥ 0.9	S2, S3, S4, S5, S7, S9, S12, S13, S15, S16, S19, S21, S25, S26, S27, S28, S29, S31, S33, S36, S37, S38, S43, S45, S48, S50B, S51, S52, S53, S54, S55, S56, S57, S58, S60, S61	36
McDonald's	$0.80 \leq \omega < 1$	S5, S8, S15, S34, S37, S40	6
Test split-half	0.80, 0.89	S23, S29	2
Compound Reliability	0.96	S48	1

When examining the table, it was found that only one article did not calculate at Cronbach's Alpha coefficient, while Cronbach's Alpha coefficient was calculated in all other articles; approximately one-tenth of them calculated McDonald's coefficient. In addition to these, test split-half was utilized in two articles, and compound reliability was examined in one article. Furthermore, it was observed that in four studies where the Cronbach's Alpha method was used (C11B, C18, C24, C35), the reliability coefficient for the entire scale was not provided.

DISCUSSION AND RESULTS

When the scales derived from scale development and adaptation studies in the field of educational sciences published in journals indexed in TR Index in 2023 were examined, it was observed that there were the highest number of studies on attitude and competence/self-efficacy scales, while studies on other topics remained limited in number. Findings regarding the abundance of attitude scales were also reached in the studies conducted by Tosun and Taşkesenligil (2014), Gül and Sözbilir (2015), Tekin and Bolat (2018), Çelik and Yüksel (2020), and Cin Şeker and Yücel Çetin (2022). The abundance of attitude scales can be explained by their wide range of applications (Cin Şeker and Yücel Çetin, 2022). Additionally, being a frequently researched variable makes the development of attitude scales in different areas important. The diversity of scale themes is viewed positively in the field of educational sciences.

When examining the EFA data, it was observed that in a large portion of adaptation studies, EFA was not conducted. Orçan (2018) stated that conducting only CFA in adaptation studies could lead to some problems. It has been stated that in case of a translation-related error, conducting only CFA may lead to a result different from what is normally expected, and the model may exhibit a wrong fit. Additionally, it has been indicated that EFA should be conducted to detect possible errors that may arise considering that the dataset may fit with multiple CFAs, and to identify cultural differences (Orçan, 2018). Contrary to the findings of our study, Boztunç, Öztürk, Eroğlu, and Kelecioğlu (2014) and Çüm and Koç (2013) determined in their researches that in scale adaptation studies, more often EFA is conducted instead of CFA. In scales where EFA is applied, the Kaiser-

Meyer-Olkin (KMO) coefficient was initially examined. The KMO coefficient provides information about whether the data in the scale are suitable for factorization, and it is expected to be greater than 0.6 (Yılmaz and Altinkurt, 2013). Field (2009) considers the KMO coefficient to be good if it falls between 0.7 and 0.8, very good if it falls between 0.8 and 0.9, and excellent if it is above 0.9. In the articles examined within the scope of our research, the KMO values ranged from 0.79 to 0.98, indicating that all studies were suitable for factorization. It is observed that rotation process is mostly used in the articles. Since rotation process facilitates the interpretation of factors, the use of these methods has been considered natural. The rotation process is applied in two ways: orthogonal rotation and oblique rotation (Gül and Sözbilir, 2015). In the assumption that factors are unrelated, orthogonal rotation is used, while in the assumption that factors are related, oblique rotation methods are used (Yurdabakan and Çüm, 2017). Among the articles examined in the study, it is observed that the Varimax method, which is one of the most used orthogonal rotation methods, is utilized the most. Büyüköztürk (2002) has stated that both orthogonal and oblique rotation methods give similar results, but due to the ease of interpretation in orthogonal rotation, this method is frequently used. However, Byrne (2001) has expressed that although interpreting orthogonal rotation methods may be easier, oblique rotation methods are more appropriate in behavioral sciences. However, Kim and Mueller (1978) did not see any drawback in using orthogonal rotation method when the purpose of the rotation process is to interpret the relationship between items and factors. Therefore, the predominant use of orthogonal rotation method in the articles included in the research has been interpreted as a common practice.

When examining the findings regarding CFA information, it was found that there was no information indicating CFA in only one study, while CFA was applied in all other studies, and fit indices were largely considered. The overall index values of the scales were generally in the "Good Fit" range; it was determined that the number of studies falling outside the ranges of Good Fit and Acceptable Fit values was very few, and these values are generally close to the "Acceptable" fit value. Researchers make decisions solely based on the p-value when conducting a χ^2 analysis or t-test, but for CFA, the decision on whether the model is consistent with the theory is based not on a single value but on the values of various fit indices. Even if the factor loadings of the items are good when the model is constructed, this fit may not be sufficient in the fit indices. Due to the wide variety of fit indices, a standard value has not been accepted (Munro, 2005; Şimşek, 2007). When the fit index values examined in the scales are compared with the literature, it has been found that the χ^2/df ratio is the mostly analyzed at fit index in researches (Munro, 2005). Each fit index has a different characteristic; for example, while the RMSEA value measures the approximate fit of the model in the population, the GFI value determines to what extent the model measures the covariance matrix in the sample. The AGFI value is used to address the issues encountered by the GFI test in high sample sizes, while CFI determines the difference between the model assumed with no relationship between variables and the null model (Çokluk, Şekercioğlu, & Büyüköztürk, 2010). Therefore, it is considered important in research to examine multiple fit indices to confirm the model.

When examining the number of modifications applied in the scales, it is observed that this variable is categorized into seven categories. While most studies do not have any modifications, this is followed by 1-2 modifications and 3-4 modifications. Gökdemir and Yılmaz (2023) stated that modifications should be made when model fit is not achieved, but the number of modifications should be limited to three. On the other hand, Gürbüz and Şahin (2018) recommend conducting EFA when structural validity and factorial validity are not achieved, even if an acceptable number of modifications are made. In this regard, it is believed that scales with more than three modifications are not considered to be reliable measures in this study.

It has been determined that all scales use the Likert type as the data collection tool. Tezbaşaran (2008) stated in his study that the Likert type is commonly used due to its usefulness and its ability to increase the level of grading. These scales are categorized into six types of Likert scales. It has been determined that the majority of the studies are prepared using a five-point Likert scale. This finding is consistent with the results obtained from the scales examined by Tekin and Bolat (2018), Çelik and Yüksel (2020), and Cin Şeker and Yücel Çetin (2022).

The scales examined within the scope of the research are categorized into five groups based on the number of items. It has been observed that the majority of the examined scales have studies with item counts of 11-20 and 21-30. No study was found in the research where the number of items was not provided. In a study examining scale development studies used in music education by Çelik and Yüksel (2020), it was also found that the most common scale studies included 11-20 and 21-30 items.

The difference between the item pool and the final number of items in the scales is categorized into eight groups. It has been observed that the most common item elimination occurs in the ranges of 0-10 and 11-20 items. Similarly, in the scale studies examined by Çelik and Yüksel (2020), it was determined that the highest number of eliminated items fell within the range of 0-10 and 11-20. Tezbaşaran (2008) stated that it is necessary to create a pool of items three to four times the number of items intended to be used in the final form. This way, there will be a higher chance of selecting items that achieve the desired discrimination. Among the examined scales, a small number of scales with item count differences of 51-60, 61-70, and 101-110 were encountered. In these scales, the number of items in the item pool exceeds three to four times the number of items in the final form. Such a large gap between the item pool and the final form may suggest that the rules of scale preparation have not been followed carefully.

The dimensions of the scales are categorized into eight groups. It has been determined that the scales with the highest number of dimensions are three-dimensional, four-dimensional, and one-dimensional. This finding is consistent with the study conducted by Çelik and Yüksel (2020).

The scales are categorized into five groups based on sample size. It has been observed that research is most commonly conducted with a sample size of 251-500 participants. When examining the sample size for EFA and CFA studies, it was determined that in sample groups of 251-500, 501-750, 751-1000, and 1001-1250 participants, different samples were mostly used for both EFA and CFA. However, in sample groups of 1-250 individuals, it was found that either separate samples were used for EFA and CFA, or only CFA was conducted. Gül and Sözbilir (2015) found that nearly half of the articles they examined had sample sizes ranging from 301 to 500, while about a quarter had sample sizes of 300 and below. Çelik and Yüksel (2020) determined that the sample sizes of the scales they examined were mostly in the ranges of 101-200 and 201-300. There are similarities between these findings and the findings of our research. Having a large sample size has a positive effect on increasing the validity and reliability of studies. In their study, Tosun and Taşkesenligil (2014) mentioned that the sample sizes commonly used in scale development studies were often between 101-200 participants. Selecting a small sample size can negatively affect the validity and reliability of scale development (Delice, 2010). When comparing the sample sizes and the number of items in the scale development and adaptation studies examined in our research, it was observed that the sample size exceeded five times the number of items for all studies. This finding is consistent with the view in the literature, which suggests applying a sample size of at least five times the number of accepted and available items (Tavşancıl, 2002).

In research, it can be observed that the reliability method commonly used is Cronbach's Alpha (α). As known, Cronbach's alpha is a single-application method and may be preferred due to requiring less effort and time compared to methods based on two applications (Gül and Sözbilir, 2015). Additionally, it can be said that it is a correct choice due to its suitability for assessing reliability in Likert-type scales (Çüm and Koç, 2013). Çüm and Koç (2013) determined that approximately 67% of the scales they examined used Cronbach's alpha as the measure of internal consistency.

According to Yang and Green (2011), the reason of the frequent use of Cronbach's alpha coefficient is its ease of interpretation and not requiring personal judgments. When the literature reviewed, it has been determined that for the Cronbach's alpha coefficient to be used, the assumption of equality of item factor loadings must be met. However, it is noted that it is not easy to ensure that all items have equal factor loadings, and in cases where factor loadings are not equal, it is more appropriate to use McDonald's omega coefficient (Yurdugül, 2006). When examining studies that

utilize both reliability methods, the close results between the two coefficients indicate that the methods used have been appropriately chosen. Cronbach's alpha coefficient distributions are categorized according to the criteria established by George and Mallery. According to these criteria, if the reliability coefficient is $\alpha < 0.5$, it is considered "Unacceptable"; in the range of $0.5 \leq \alpha < 0.6$, it is "Weak"; in the range of $0.6 \leq \alpha < 0.7$, it is "Acceptable"; in the range of $0.7 \leq \alpha < 0.9$, it is "Good"; and if $\alpha \geq 0.9$, it is interpreted as "Excellent" (George and Mallery, 2003). Among the examined studies, it is observed that 36 studies can be interpreted as "Excellent", 26 as "Good", and 2 as "Acceptable" based on Cronbach's Alpha coefficient. Additionally, according to the classification determined by Özdamar (2002), if the reliability coefficient is between $0.61 < \alpha < 0.80$, the scale is considered to have moderate reliability, and if it is between $0.81 < \alpha < 1.00$, it is considered to have high reliability. According to this classification, all the examined scales do not pose a problem in terms of reliability. In the examined studies, scales in S11B, S18, S24, S35, and S59 did not provide an overall reliability coefficient for the scale, but they only provided reliability coefficients for their dimensions. Additionally, in some studies, multiple reliability methods have been examined, aiming to prevent possible errors by using multiple reliability methods.

Based on the results obtained from the study, several recommendations have been proposed for future scale development or adaptation studies. These recommendations are as follows:

A comprehensive guidebook should be published for future scale development and adaptation studies, and researchers should meticulously adhere to the stages outlined in this guidebook during the process of scale development or adaptation. A center should be established in our country for the preparation and control of scale development and adaptation studies. Researchers intending to develop or adapt scales should not repeat the steps followed in previous studies without thorough investigation. It should be remembered that repeating a mistake can lead to even greater problems. Researchers should have detailed knowledge of different validity and reliability methods and should use the most suitable method for their research. Courses related to scale development and adaptation should be added to postgraduate education to enable researchers to obtain accurate information from reliable sources.

Conflicts of Interest: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding Details: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRedit Author Statement: Müge Vural: Conceptualization, Methodology, Writing-Original draft preparation. Murat Tuncer: Supervision, Data curation, Writing- Original draft preparation. Ahmet Egemen Akmençe: Writing- Reviewing and Editing, Visualization, Investigation, Validation.

Ethical Statement: This study does not require ethics committee approval since it did not involve collecting data from Human participants.

REFERENCES

- Acar Güvendir, M. ve Özer Özkan, Y. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi. *Elektronik Sosyal Bilimler Dergisi*, 14(52), 23-33.
- Arıkan, R. (2013). *Araştırma yöntem ve teknikleri*. Ankara: Nobel Akademik Yayıncılık.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.

- Boztunç Öztürk, N., Eroğlu, M. G. ve Kelecioğlu, H. (2014). *Eğitim bilimleri alanında yapılan ölçek uyarlama makalelerinin incelenmesi*. IV. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, Ankara, 9-13 Haziran 2014.
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi*, 32, 470-483.
- Byrne, B.M. (2001). Structural equation modeling with AMOS: Basic concepts, applications and programming. *Mahwah, NJ: Lawrence Erlbaum Associates*, 92-93.
- Cin Şeker, Z. ve Yücel Çetin, D. (2022). Okuma eğitimi alanında yapılan ölçek geliştirme çalışmalarının incelenmesi: Bir meta-sentez çalışması. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi*, 11(3), 1193-1209.
- Cohen R.J. ve Swerdlik M.E. (2010). *Psychological testing and assessment*. Boston: McGraw-Hill Companies.
- Çalık, M. ve Sözbilir, M. (2014). İçerik analizinin parametreleri. *Eğitim ve Bilim*, 39(174), 33-38.
- Çelik, D. ve Yüksel, G. (2020). Müzik eğitimi kapsamında yapılan ölçek geliştirme çalışmalarının çok yönlü incelenmesi. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, 9(5), 4059-4087.
- Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları*. Ankara: Pegem.Net, p.275-85.
- Çüm, S. ve Koç, N. (2013). Türkiye’de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *Eğitim Bilimleri ve Uygulama*, 12(24), 115-135.
- Delice, A. (2010). Nicel araştırmalarda örneklem sorunu. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(4), 1968-2018.
- Delice, A. ve Ergene, Ö. (2015). Ölçek geliştirme ve uyarlama çalışmalarının incelenmesi: Matematik eğitimi makaleleri örneği. *Karaelmas Journal of Educational Sciences*, 3(1), 60-75
- Deniz, Z. (2007). The adaptation of psychological scales. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 40(1), 1-16.
- Dirlik, E. (2013). *Eğitim kurumlarında kullanılan psikolojik testlerin ölçme standartlarına göre incelenmesi* [Yayımlanmamış yüksek lisans tezi]. Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Edenborough R. (1999). *Using psychometrics: a practical guide to testing and assessment*. London: Kogan Page.
- Erkorkmaz, Ü., Etikan, İ., Demir, O., Özdamar, K. ve Sanisoğlu, S. Y. (2013). Doğrulayıcı faktör analizi ve uyum indeksleri. *Türkiye Klinikleri Journal of Medical Sciences*, 33(1), 210-223.
- Erkuş, A. (2019). *Psikolojide ölçme ve ölçek geliştirme I: Temel kavramlar ve işlemler (4. Baskı)*. Pegem Akademi Yayıncılık.
- Field, A. (2009). *Discovering statistics using SPSS*. London: SAGE Publications Ltd.
- George, D. ve Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference. 11.0 update (4th ed.)*. Boston: Allyn & Bacon.

- Gökdemir, F. ve Yılmaz, T. (2023). Processes of using, modifying, adapting and developing likert type scales. *J Nursology*, 26(2), 148-160.
- Gül, Ş. ve Sözbilir, M. (2015). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi. *Eğitim ve Bilim*, 40(178), 85-102.
- Güler, N. (2011). *Eğitimde ölçme ve değerlendirme*. Ankara: PegemA.
- Gürbüz, S. ve Şahin, F. (2018). *Sosyal Bilimlerde Araştırma Yöntemleri Felsefe Yöntem-Analiz*. (5. Baskı). Ankara: Seçkin Yayıncılık.
- Karagöz, Y. (2017). *SPSS ve AMOS uygulamalı nitel-nicel-karma bilimsel araştırma yöntemleri ve yayın etiği*. Ankara: Nobel Yayınevi.
- Kim, J.O. ve Mueller, C.W. (1978). Introduction to factor analysis: What it is and how do it. *Beverly Hills, CA: Sage Publications*, 54.
- Mor Dirlik, E. (2014). Ölçek geliştirme konulu doktora tezlerinin test ve ölçek geliştirme standartlarına uygunluğunun incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(2), 62-78.
- Munro, B.H. (2005). *Statistical methods for health care research*. Philadelphia: Lippincott Williams & Wilkins, p.351-76.
- Olgun, G. ve Alatlı, B. (2019). Türkiye’de ergenlere yönelik ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *TEBD*, 19(1), 568-592.
- Orçan, F. (2018). Açımlayıcı ve doğrulayıcı faktör analizi: İlk hangisi kullanılmalı? *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi (Journal of Measurement and Evaluation in Education and Psychology)*, 9(4);413-421
- Özdamar K. (2002). *Paket programlarla istatistiksel veri analizi-1 (4. Baskı)*. Eskişehir: Kaan Kitabevi.
- Özgüven, İ. E. (2011). *Psikolojik testler*. Ankara: Pdrem Yayınları.
- Şahin, N. (1994). Psikoloji araştırmalarında ölçek kullanımı. *Türk Psikoloji Dergisi*, 9(33), 19-26.
- Şahin, M. G. ve Boztunç Öztürk, N. (2018). Eğitim alanında ölçek geliştirme süreci: Bir içerik analizi çalışması. *Kastamonu Eğitim Dergisi*, 26(1), 191-199.
- Şekercioğlu, G. (2023). *Psikolojik testler* (Ed. Yasemin Yavuzer, Zeynep Karataş). Pegem Akademi Yayıncılık.
- Şimşek, Ö. F. (2007). *Yapısal eşitlik modellemesine giriş, temel ilkeler ve LISREL uygulamaları*. Ankara: Ekinoks, p.4-22.
- Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel Yayıncılık.
- Tavşancıl, E. ve Aslan, E. (2001). *Sözel, yazılı ve diğer materyaller için içerik analizi ve uygulama örnekleri*. İstanbul: Epsilon Yayıncılık.
- Tekin, M. T. ve Bolat, Y. (2018). Türkiye’de yazma eğitimi üzerine yapılan ölçek geliştirme çalışmalarının analizi: Bir meta-sentez çalışması. *Hatay Mustafa Kemal Üniversitesi Eğitim Fakültesi Dergisi*, 2(2), 26-40.

- Tezbaşaran, A. (2008). *Likert tipi ölçek hazırlama kılavuzu*. Ankara: Türk Psikologlar Derneği Yayınları
- Tosun, C. ve Taşkesenligil, Y. (2014). *Türkiye’de fen eğitimi alanında geliştirilen/adapte edilen ölçeklerin ve başarı testlerinin doküman analizi*. XI. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi, Adana.
- World Health Organization (WHO). (2017). Process of translation and adaptation of instruments.
- Yang, Y. ve Green, S.B. (2011). Coefficient alpha a reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, 29(4), 377-392.
- Yıldırım, A. ve Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.
- Yılmaz, K. ve Altınkurt, Y. (2013). Örgütsel iklim ölçeğinin türkçeye uyarlanması: Geçerlik ve güvenilirlik çalışması. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, 3(1), 1-11
- Yurdabakan, İ. ve Çüm, S. (2017). Davranış bilimlerinde ölçek geliştirme (Açıklayıcı faktör analizine dayalı). *TJFMPC*, 11(2): 108-126.