



(ISSN: 2602-4047)

Yılmaz, A., & Özgül, İ. (2023). An Examination of Scale Development Studies in the Field of Educational Sciences Within the Scope of International Standards, *International Journal of Eurasian Education and Culture*, 8(24), 2895-2920.

DOI: <http://dx.doi.org/10.35826/ijoec.1810>

Article Type (Makale Türü): Research Article

AN EXAMINATION OF SCALE DEVELOPMENT STUDIES IN THE FIELD OF EDUCATIONAL SCIENCES WITHIN THE SCOPE OF INTERNATIONAL STANDARDS

Adem YILMAZ

Assoc. Prof. Dr., Kastamonu University, Kastamonu, Türkiye, yilmazadem@kastamonu.edu.tr
ORCID: 0000-0002-1424-8934

İlhan ÖZGÜL

Assoc. Prof. Dr., Kastamonu University, Kastamonu, Türkiye, ilhanozgul@kastamonu.edu.tr
ORCID: 0000-0002-5615-7934

Received: 11.05.2023

Accepted: 14.11.2023

Published: 04.12.2023

ABSTRACT

This study aimed to examine the scale development studies conducted in the field of educational sciences within the scope of international standards. The document analysis method, one of the qualitative research approaches, was used in the study. In the research process, the scale development studies in the field of educational sciences originating from Türkiye were analysed. A total of 85 studies carried out between the years 2010 and 2023 were included in the research. Within the scope of the research, the criteria such as subject area, sample size, sample type, validity and reliability analyses, item and factor elimination criteria, statistical software were utilized, and researcher status were considered. As a result of the research, it was determined that most scale development/adaptation studies did not meet international standards or were able to meet the standards solely at a certain level. In terms of the sample type variable, there exist quite different types of studies. When the sample size results were reviewed, it was discovered that the samples were insufficient. It was also unearthed that validity and reliability analyses were employed to a great extent but were still insufficient. It was found out that limited or no information about item and factor elimination was given. Whilst statistical software is resorted to in all the studies, some aspects such as researcher status are not included much. As a result of the research, various suggestions were made.

Keywords: Educational sciences, scale development, international standards.

INTRODUCTION

A fair number of studies are conducted to provide quality education and training services. These studies vary in individual, organizational, societal, and statistical aspects. One of the areas where education services are intensively provided is the field of higher education. In higher education, some studies can give direction to society, have scientific foundations, and are carried out by academically competent researchers. Checking these studies from time to time, analyzing them from different perspectives and examining their qualities can help provide a better quality education service and contribute to the systematic improvement of scientific developments (Mor-Dirlik, 2014). In the light of scientific findings obtained in certain fields and levels of education and the accumulation of scientific knowledge, new methods, tools and techniques developed in the field of measurement have become widespread in other fields and levels of education (Uçan, 2005).

Learners' cognitive, affective and psychological characteristics are measured at almost every level, from primary school to tertiary level (Çüm, 2018; Yılmaz, 2021). While performing these measurement procedures, students are usually compared with a test and their responses to the questions are scored using a particular method (Anastasi, 1988). Amongst these tests, psychological tests, scales, and questionnaires are frequently referred to. Because such tests can warrant objective results, have high validity and reliability and can be easily applied (Conway, 2006). With a view to determining the characteristics of individuals, a number of measurements should be realized first. These measurements should be performed using tools whose validity and reliability are confirmed in the relevant line of literature. Although psychological tests are used extensively in higher education, scales are the most preferred data collection tool. Researchers may need many different types of scales in their studies though. In this regard, they either use ready-made scales in the literature or develop or adapt scales according to their needs. However, this is not always executed in a healthy fashion since scales can serve their purpose only if they are systematic and can meet specific standards (Özoğlu, 2007). In other words, for the information obtained to be meaningful and usable, these scales should be created meticulously, in accordance with the set standards going through certain stages (Baykul & Turgut, 2010).

In the sphere of educational sciences, there has been an intensive increase in scale development studies in recent years. This is frequently due to meeting the need for scientific research and to the wish to designate novel and original forms of research in the area. Researchers who prefer scale development should own the in-depth knowledge of the content and structure of the feature they want to measure. In doing so, they need to examine the previous scale development studies and complete this process in specific steps. That said, scale development studies are long-term, and they require gained experience of the related individuals. At the same time, scholars who intend to develop scales should have done previous studies in this field, have a good command of the literature, and possess sufficient statistical knowledge (Gül & Sözbilir, 2015). In addition, it is essential to revise the developed scales depending on new developments, adjust them through different samples, check their psychometric measurements and renew them (Azaltun, 2008). In the accumulated literature, there are plentiful scale development studies and studies that give direction to these bulk of studies

(Ayyıldız & Yılmaz, 2021; Cabrera-Nyugen, 2010; Carpenter, 2018; DeVellis & Thorpe, 2021; DiStefano, Zhu, & Mindrila, 2009; Erkuş, 2012; Knafl et al., 2012; Simms, 2008). Nevertheless, the fact that there is publication pressure on academics, scale development studies is considered an easy way out in the field, leading to an increase in these studies and bringing many errors and deficiencies (Erkuş, 2007).

Researchers may sometimes prefer scale adaptations instead of the endeavor toward scale development (Yılmaz, 2018). This is owing to the fact that scale adaptation is both easier and faster than scale development (Hambleton & Patsula, 1999). In the territories of educational sciences, there are many independent studies in which different scale development processes are adopted. Thereupon, these differences are moulded over time and can be accepted as 'unwritten rules'. However, a systematic and acceptable method should always be followed in scale development/adaptation studies (Acar-Güvendir & Özer-Özkan, 2015). To be able to create a new scale every day and interpret the mass of information meaningfully, reliable, valid and comprehensive studies are needed (Akgöz, Ercan & Kan, 2004). When the literature is delved into, it is clear that meta-studies are conducted from time to time. Çüm (2013) analysed the article studies conducted in the field of education science and psychology. Mor-Dirlik (2014) observed five different theses prepared within the scope of educational sciences from different perspectives. Gül and Sözbilir (2015) examined scale development studies in the field of science and mathematics thematically. Tosun and Taşkesenligil (2014) dwelled on the studies developed within the scope of science education. Boztunç-Öztürk, Eroğlu, and Kelecioğlu (2015) investigated the scale adaptation articles in the field of education. Güler and Ayan (2020) examined 112 attitude scale development studies conducted in the field of educational sciences in Türkiye between 2002-2018 and investigated to what extent the process steps that should be taken into account in the development of attitude scales were met. As it is seen, it is deemed necessary to repeat these studies frequently. For this very reason, re-examining the scale development studies in educational sciences serves for the literature. The main purpose of this research is then to inspect the scale development studies conducted in the field of educational sciences within the scope of international standards. In this frame of reference, answers to the following sub-problems were sought:

1. What subject areas are preferred in scale development?
2. What kind of differences do scale development studies create as regards sample type and sample size?
3. What kind of validity analyses were deployed in scale development studies?
4. What kind of reliability analyses were used in scale development studies?
5. What are the item elimination criteria found in the scale development process?
6. What are the factor elimination criteria found in the scale development process?
7. What kind of statistical software was evident in scale development?
8. What level of information was given about researcher status in scale development studies?

METHOD

In this study, the scale development studies conducted in Türkiye between the years 2010 and 2023 were analysed with the help of the content analysis method. Content analysis studies are frequently turned to so as to analyse concepts and phenomena with the help of documents (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz & Demirel, 2008). Content analysis can generally be categorised under three different headings. These titles can be expressed as meta-analysis, meta-synthesis, and descriptive content analysis (Çalık & Sözbilir, 2014). With this research, scale development studies in the field of educational sciences were explored respecting certain variables and compared within the scope of international standards. In light of these, the present research can be expressed as descriptive content analysis.

Scale Development Studies Subjected to Content Analysis

The scale development studies subjected to content analysis were conducted between the years 2010 and 2023. In the current research, it is noteworthy to articulate herein that merely the scale development studies in the educational sciences field were concentrated on.

Data Collection Process

The scale development studies obtained in the research process consist of studies conducted in Türkiye between the years 2010 and 2023. While searching for the said studies, varying keywords such as "educational sciences, scale development, validity and reliability study" were recruited. At the same time, the publications in field indexes such as TRDizin-ULAKBIM, ERIC, H.W. Wilson and SSCI level indexes were gone through.

The reason for examining these databases is that the studies published in these fields already have a certain standard and are subject to evaluation according to a set of specific criteria. In the research process, some criteria were also considered that were apparent in the journals in which scale development articles were published. These criteria are dependent on these factors:

- High impact factors of journals,
- The selected journals are of high quality in accordance with TUBITAK publication incentive criteria,
- Having internationally recognised indexes,
- Journals publishing in the field of educational sciences,
- Regular publication in journals between the years 2010 and 2023,
- Journal editors' and field editors' being competent in the field of assessment and evaluation,
- The journals' having a long and well-established publication history.

Data Analysis

The data obtained within the scope of the research were subjected to content analysis, and categorical classification and frequency analysis were presented. In this direction, the application data were analysed

descriptively and visualised with the help of frequency and percentage ratios. Further, the results obtained by the researchers in the studies were coded and categorised.

Compliance with Ethical Rules and Obtaining the Necessary Permissions

In the research process, the maximum attention was paid to ethical rules, and the necessary ethics committee permissions were obtained accordingly. The ethics committee permission for the research was approved by Kastamonu University Social and Human Sciences Research and Publication and Ethics Committee with the decision dated 04.05.2020 and issued 30.

FINDINGS

The research results were analysed separately for each sub-problem situation. The first problem statement of the research is "What subject areas are preferred in scale development?". Figure 1 shows the preferred subject areas in this sense.

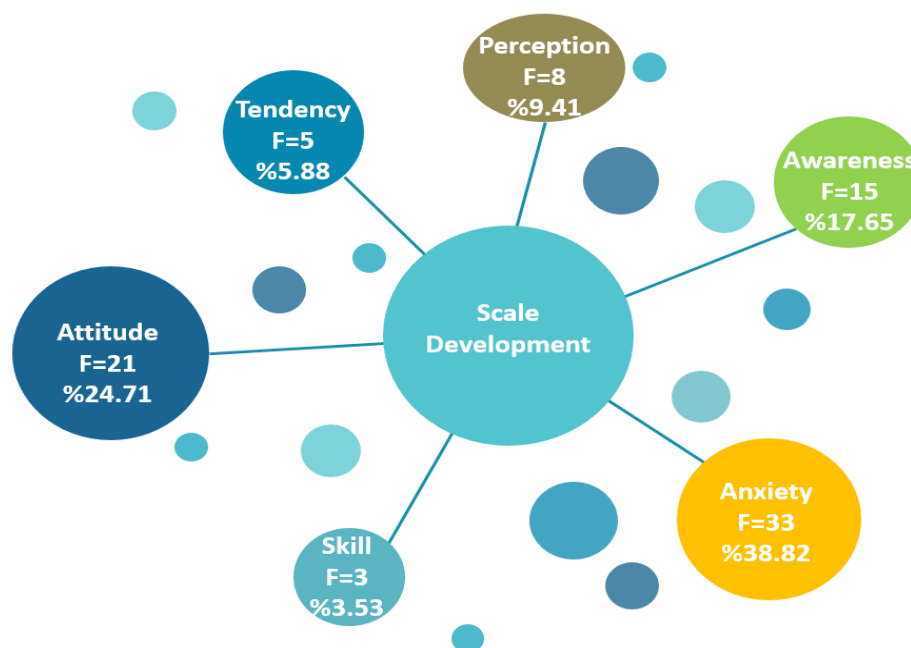


Figure 1. Scale Development Subject Areas

When Figure 1 is checked, it seems that scale development studies are divided into anxiety, attitude, awareness, perception, tendency, and skill according to the subject areas. Whereas the most developed scale type according to the subject areas is the anxiety scale, the least developed scale type is skills. The second problem statement of the research is "What kind of differences do scale development studies create as regards sample type and sample size?". Table 1 displays the findings regarding sample type and sample size.

Table 1. Findings on Sample Type and Sample Size

Sample Type	(f)	%
Preschool level	3	3.53
Primary education level	18	21.18
Secondary education level	28	32.94
Undergraduate level	22	25.88
Postgraduate level	14	16.47
Total	85	100

Sample Size	(f)	%
Between 0-100	2	2.35
Between 101-150	7	8.24
Between 151-250	11	12.94
Between 251-400	16	18.82
Between 401-600	17	20.00
Between 601-800	13	15.29
Between 801-1000	11	12.94
1000 and above	8	9.41
Total	85	100

Table 1 shows that scale development studies were preferred at preschool, primary, secondary, undergraduate and graduate levels according to the sample type. Among the sample types, the secondary education level was preferred the most, and the preschool level was preferred the least. When the scale development studies were focused on according to the sample size, it is understood that many different sample sizes were chosen. Among the sample sizes in question, the most preferred size was between 401-600, while the least preferred size was between 0-100. The third problem statement of the research is *"What kind of validity analyses were deployed in scale development studies?"*. Table 2 exhibits the findings related to validity analyses.

Table 2. Findings for Validity Analysis

Validity Analysis	(f)	%
Content validity	15	17.65
Face validity	11	12.94
Criterion validity	22	25.88
Construct validity	37	43.53
Total	85	100

Taking a glance at Table 2 one sees that content, face, criterion, and construct validities were analysed for validity analysis in scale development studies. Among these analyses, construct validity was analysed the most and face validity was analysed the least. Criterion validity and content validity are other preferred types of validity analyses. The fourth problem statement of the research is *"What kind of reliability analyses were used in scale development studies?"*. Table 3 gives the findings related to reliability analyses.

Table 3. Findings for Reliability Analysis

Reliability Analysis	(f)	%
Test-retest	22	25.88
Equivalent forms	11	12.94
Internal consistency	36	42.35
Split-test	12	14.12
Sub-under group analysis	4	4.71
Total	85	100

Looking at Table 3 one can arrive at that test-retest, equivalent forms, internal consistency, split-test and sub-under group analyses were performed for reliability analysis in scale development studies. Among these analyses, internal consistency was the most frequently preferred one and sub-under group analysis was the least preferred one. Even though internal consistency analyses are classified as dissimilar subtypes, it is the most frequently preferred type of analysis. The fifth problem statement of the research is "What are the item elimination criteria found in the scale development process?". Table 4 demonstrates the findings respecting the item elimination criteria.

Table 4. Findings on Item Elimination Criteria

Item Elimination Criteria	(f)	%
Item factor loads	29	34.12
Overlapping loads	14	16.47
Item analysis	22	25.88
Expert opinion	7	8.24
Item-total correlation	8	9.41
Common variances	5	5.88
Total	85	100

Table 4 unveils that item factor loadings, overlapping loadings, item analysis, expert opinion, item-total correlation and common variances are used as item elimination criteria in scale development studies. Among these criteria, item factor loads are the most preferred and common variances are the least preferred. The sixth problem statement of the research is "What are the factor elimination criteria found in the scale development process?". Table 5 presents the findings regarding the factor elimination criteria.

Table 5. Findings on Factor Elimination Criteria

Factor Elimination Criteria	(f)	%
Eigenvalue results	41	48.24
Scree plot graph	23	27.06
Variance ratio	15	17.65
Number of items	2	2.35
Expert opinion	2	2.35
Results of the literature	2	2.35
Total	85	100

When Table 5 is analysed, one contemplates that eigenvalue results, scree plot graph, variance ratio, number of items, expert opinion and literature results are preferred for factor elimination criteria in scale development

studies. Among these, the most preferred item was the eigenvalue results. The least preferred criteria were the number of items, expert opinion, and literature results. The seventh problem statement of the research is "What kind of statistical software was evident in scale development?". Figure 2 portrays the findings in relation to the statistical software used in the scale development process.

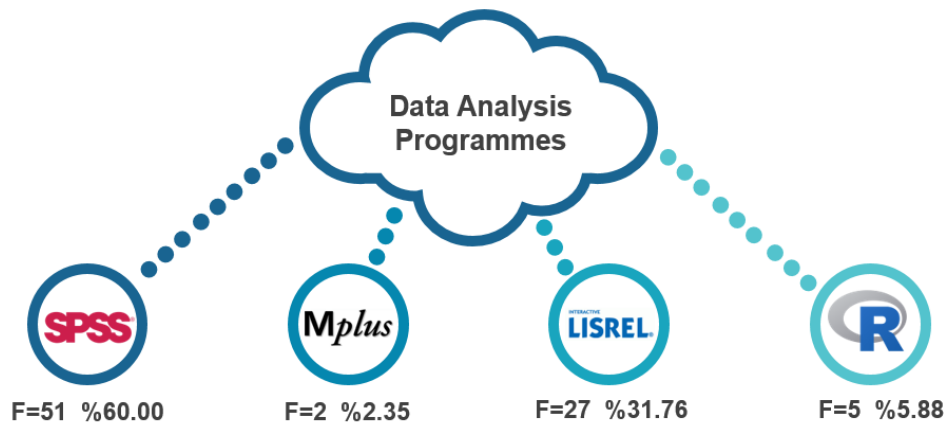


Figure 2. Statistical Software Used in Scale Development Process

When Figure 2 is revised, it is clear cut that SPSS, Mplus, LISREL and R programmes are preferred according to the statistical software used in the scale development process. Among these software programmes, SPSS was the most preferred program. *This was followed by LISREL, R and Mplus programmes. The eighth problem statement of the research is "What level of information was given about researcher status in scale development studies?".* Table 6 lays out the findings on researcher status.

Table 6. Findings Related to Researcher Status

Researcher Status	(f)	%
Master's Degree Student	23	27.06
PhD Student	26	30.59
Academician	34	40.00
Unspecified	2	2.35
Total	85	100

Table 6 depicts that the researchers in the scale development studies were mostly academicians and postgraduate students. In 2 studies, researcher status was not specified.

CONCLUSION, DISCUSSION and RECOMMENDATIONS

In this study, in which the scale development studies conducted in the field of educational sciences were evaluated within the framework of international standards, the subject areas were first analysed. When 85 scale development studies were gone through in this respect, it was figured out that the studies were intensively prepared on anxiety, attitude and awareness. In fact, this has been emphasised in many studies in the literature (Gül & Sözbilir, 2015; Tosun & Taşkesenligil, 2014). Critical results can be obtained because

anxiety, attitude and awareness are the behaviours that individuals reflect the most easily. Similarly, in another study conducted by Çiltaş, Güler, and Sözbilir (2012), it is declared that attitude and anxiety studies are frequently preferred in scale development studies. The least common scale type in scale development studies is skill scales. This is on account of the fact that skills can generally be observed through long-term observations and regular measurements made in a certain period of time. Hence, the scarcity of skill scales can be considered a normal situation.

The second variable in scale development studies is the sample type and size. When the sample type is considered, it emerges that studies have been conducted at many different levels starting from preschool period to higher education. Among these studies, studies at the preschool level are pretty few. This may be because it is oftentimes difficult for preschool children to respond to scale questions. That being said, the inability to pay attention of the children as participants of research for a long time might be another reason. As a sample type, most of the studies were planned at the secondary education level. This may be because the high school level is a transition stage, and students have both vocational concerns and high expectations for the future prior to moving on to higher education (Gençtürk & Memiş, 2010). These results are supported by the study conducted by Tosun and Taşkesenligil (2014). Undergraduate students constitute another group with the highest number of studies according to the sample type. The main reason for this situation is that undergraduate students form an easily accessible sample group. At the same time, reaching a relatively higher number of people in a quick sense is another prominent feature of this group (Deveci, Çalmaz & Açıık, 2012). There is no international scale development standard as to subject area or sample type variables. Albeit, when the issue is handled in respect of psychological test measurements, the diversity and richness of the subject area and sample type can be thought as features that strengthen the overall profile of the existing literature.

When the scale development studies are observed apropos of sample size, the most dense sample group is between 401-600 persons and the number of 1000 and more people is at low levels. There are some standards in the literature on sample size. These standards are firstly related to the number of items. Considering the international scale development standards, it is stated that the sample size should be at least five times to 10 times the number of questions (Özdamar, 2003). This is often tried to be kept high, mainly because the statistical programmes run are sensitive to the sample size. On the other hand, the sample size does not change the results after a certain point. Moreover, as the sample size boosts, unrelated parameters start to show a relationship, which may result in the scale to go beyond its purpose (Yılmaz, 2018). Another rule for sample size in the scale development process was pronounced by Comrey and Lee (1992). Here, provided the sample size is 100 people, it is defined as poor, 200 people as average, 300 people as good, 500 people as very good and 1000 people as excellent. It is worth recalling at this juncture the number of items in the scale and the data saturation limit. After data saturation occurs, elevating the sample size will not alter the results much (Delice, 2010). Considering the studies addressed within the scope of the research, it was identified that the sample size of a considerable number of studies was insufficient thinking about the number of questions, and the value ranges accepted in the scales were not good enough.

When the research results are probed within the scope of validity analyses, it is realized that the most preferred type of validity is construct validity. Especially in the scale development studies conducted after the 2000s, exploratory and confirmatory factor analyses have started to materialise in scores of studies with the help of computer technologies (Erkuş, 2012; Yurdugül, 2005). Since the necessity of confirmatory factor analyses is reiterated in the studies of the literature, it can be viewed as a natural outcome where the controls for construct validity are intensive (Brinkmann, 2009). This is followed by criterion validity and content validity. As a matter of fact, the need for content validity has been punctuated in the studies in the literature (Slavec & Drnovsek, 2012). When validity analyses are scanned by considering international standards, it is articulated that content, face and construct validity are vital (Yurdugül & Bayrak, 2012).

When the results obtained within the scope of reliability analysis are scrutinised, it is noticed that the internal consistency coefficient is calculated in less than half of the studies. Apart from this, the test-retest method and split-test method are frequently used. In the journals that are eminent with specific standards, the calculation of the internal consistency coefficient is mostly required (Topu, Baydaş, Turan & Göktaş, 2013). For this reason, it can be put forth that the analysed studies do not meet the sufficient criteria for reliability analysis.

When the item elimination criteria were prospected within the scope of the research, it was seen that the most frequently preferred method was the calculation of item factor loadings and item analysis options. In the study conducted by Worthington and Whittaker (2006), it is announced that item factor loadings are frequently used. At the same time, another factor that should be taken into consideration within the scope of international standards is item analysis. With item analysis, scales can be better divided into sub-dimensions (Tavşancıl, 2002). Within the scope of the studies examined, the use of item factor loadings and item analysis is relatively low. When the factor elimination criteria were assayed in the study, the eigenvalue results were the most frequently used method. The scree plot graph follows this. The literature frequently emphasises the examination of eigenvalue results and scree plot graphs. In the study by Hinkin (1998), it is put forward that these values should be examined within the scope of international standards. At the same time, Erkuş (2012) italicises in his study that eigenvalue results alone are insufficient and that the results of scree plot graphs should also be evaluated. In light of these results, it would be fair to state that the analysed studies are partially sufficient in terms of eigenvalue results. Intensive statistical operations are performed in the scale development process. There are many software programmes in the literature on this subject. With this being said, SPSS and LISREL programmes are among the most preferred software. Besides these programs, software such as Mplus and R are also preferred. As a result of the research, the most preferred software was SPSS. The reason for this may be that the user interface of the SPSS programme is highly accessible whereas in LISREL, Mplus and R programs, the user interface requires higher levels of computer literacy and coding skills. Thusly, it is normal that the SPSS programme is preferred intensively.

The last variable of the research process is researcher status. Scale development studies often call for intensive labour, experience and adequate levels of field knowledge. When the research results are surveyed, it is seen

that most of the studies were developed by researchers, masters and doctoral students. At this point, the high number of master's and doctoral students may cause some concern among stakeholders. Because scale development by researchers who are not yet fully trained may still contribute to the field, yet may also lead to the production of faulty scales.

The following suggestions can be made in compliance with the results of this study:

1. Before developing a new scale in scale development studies, care should be taken to scan the literature well and use scales suitable for the purpose. A new scale development process should not be attempted unless it is really necessary.
2. Scale development studies should be carried out in areas where the literature is insufficient in sample type, and richness should be created.
3. Sample size is an essential standard in scale development studies. Thus, statistical techniques should be used to attain a sufficient sample size.
4. Content and construct validity are among the indispensable criteria in the research process. Care should be taken to use these validity types.
5. Experienced people with field knowledge should carry out the scale development process. Otherwise, many similar scales are developed, and many scales serving the same purpose are formed in the literature.

ETHICAL TEXT

This article complies with the journal's writing rules, publication principles, research and publication ethics rules, and journal ethics rules. The responsibility for any violations that may arise regarding the article belongs to the authors. The ethics committee permission of the article was obtained by Kastamonu University/Publication Ethics Committee with the decision numbered 2020/30 dated 04.05.2020.

Author(s) Contribution Rate: In this study, the contribution rate of the first author is 50%, the contribution rate of the second author is 50%.

REFERENCES

- Acar-Güvendir, M., & 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.
- Akgöz, S., Ercan, İ., & Kan, İ. (2004). Meta-analizi. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 30(2), 107-112.
- Anastasi, A. (1988). *Psychological testing*. (6th Ed.). MacMillan Publishing Co. Inc.
- Ayyıldız, P., & Yılmaz, A. (2021). 'Moving the kaleidoscope' to see the effect of creative personality traits on creative thinking dispositions of pre-service teachers: The mediating effect of creative learning environments and teachers' creativity fostering behavior. *Thinking Skills and Creativity*, 41, 100879, 1-10. <https://doi.org/10.1016/j.tsc.2021.100879>
- Azaltun, M. (2008). VI. araştırma yöntemleri semineri-ölçme ve ölçek geliştirme. *Anatolia: Turizm Araştırmaları Dergisi*, 19(1), 104-111.

- Baykul, Y., & Turgut, F. (2010). *Eğitimde ölçme ve değerlendirme*. Pegem Akademi Yayıncılık.
- Boztunç-Öztürk, N., Eroğlu, M.G., & Kelecioğlu, H. (2015). Eğitim alanında yapılan ölçek uyarlama makalelerinin incelenmesi. *Eğitim ve Bilim*, 40(178), 123-137. <https://doi.org/10.15390/EB.2015.4091>
- Brinkman, W. P. (2009). Design of a questionnaire instrument. S. Love (Edt). *Handbook of mobile technology research methods* içinde (ss.31-53). Nova Publisher.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2008). *Bilimsel araştırma yöntemleri*. Pegem Akademi Yayınları.
- 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.
- Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, 12(1), 25-44. <https://doi.org/10.1080/19312458.2017.1396583>
- Comrey, A. L., & Lee, H. L. (1992). *A first course in factor analysis*, Hillsdale, Erlbaum.
- Conway, M. J. (2006). *How to collect data: measurement and evaluation*. American Society for Training and Development.
- Çalık, M., & Sözbilir, M. (2014). İçerik analizinin parametreleri. *Eğitim ve Bilim*, 39(174), 33-38. <https://doi.org/10.15390/EB.2014.3412>
- Çiltaş, A., Güler, G., & Sözbilir, M. (2012). Türkiye'de matematik eğitimi araştırmaları: İçerik analizi çalışması. *Kuram ve Uygulamada Eğitim Bilimleri*, 12(1), 515-580.
- Çüm. S. (2013). *Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi*. [Yayımlanmamış Yüksek Lisans Tezi]. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Delice, A. (2010). Nicel araştırmalarda örneklem sorunu. *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. *Dicle Tıp Dergisi*, 39(2), 189-196.
- DeVellis, R.F., & Thorpe, C.T. (2021). *Scale development: Theory and applications*. Sage publications.
- 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.
- Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar. *Türk Psikoloji Bülteni*, 13(40), 17-25.
- Erkuş, A. (2012). *Psikolojide ölçek ve ölçek geliştirme-I: Temel kavramlar ve işlemler* (1.Baskı). Pegem Akademi Yayıncılık.
- Gençtürk, A., & Memiş, A. (2010). İlköğretim okulu öğretmenlerinin öz-yeterlik algıları ve iş doyumlarının demografik faktörler açısından incelenmesi. *İlköğretim Online*, 9(3), 1037-1054.
- Gül, Ş., & Sözbilir, M. (2015). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi. *Eğitim ve Bilim*, 40(178), 85-102. <https://doi.org/10.15390/EB.2015.4070>

- Güler, G., & Ayan, C. (2020). Review of attitude scales developed in Turkey between 2002-2018 regarding the scale development process. *Journal of Faculty of Educational Sciences*, 53(3), 839-863. <https://doi.org/10.30964/auebfd.658488>
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104-121. <https://doi.org/10.1177/109442819800100106>
- Knafl, G. J., Dixon, J.K., O'Malley, J.P., Grey, M., Deatrick, J.A., Gallo, A., & Knafl, K.A. (2012). Scale development based on likelihood cross-validation. *Statistical Methods in Medical Research*, 21(6), 599-619.
- 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.
- Özdamar, K. (2003). *Modern bilimsel araştırma yöntemleri*. Kaan Kitapevi.
- Özoğlu, S. Ç. (2007). *Eğitimde rehberlik ve psikolojik danışma*. (3.Baskı). Ankara Üniversitesi Eğitim Bilimleri Yayınları.
- Simms, L.J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, 2(1), 414-433.
- Slavec, A., & Drnovsek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review*, 14(1), 39-62.
- Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Nobel Yayınevi.
- Topu, F. B., Baydaş, Ö., Turan, Z., & Göktepe, Y. (2013). Öğretim teknolojisi araştırmalarında geçerlik ve güvenirlik önlemleri. *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*. XI. Ulusal Fen bilimleri ve Matematik Eğitimi Kongresi'nde sunulan sözlü bildiri, Adana.
- Uçan, A. (2005). *Müzik eğitimi temel kavramlar-ilkeler-yaklaşımlar ve Türkiye'deki durum* (3.Baskı). Evrensel Müzikevi.
- 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.
- Yılmaz, A. (2018). *Fen bilgisi öğretmen yetiştirme programlarında kalite standartlarının belirlenmesi: Ölçek geliştirme ve uygulama çalışması*. [Doktora Tezi]. Kastamonu Üniversitesi Fen Bilimleri Enstitüsü, Kastamonu.
- Yılmaz, A. (2021). The effect of technology integration in education on prospective teachers' critical and creative thinking, multidimensional 21st century skills and academic achievements. *Participatory Educational Research*, 8(2), 163-199. <http://dx.doi.org/10.17275/per.21.35.8.2>
- Yurdugül, H. (2005). *Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması*. 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. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, Özel Sayı 2*, 264-271.

EĞİTİM BİLİMLERİ ALANINDA ÖLÇEK GELİŞTİRME ÇALIŞMALARININ ULUSLARARASI STANDARTLAR KAPSAMINDA İNCELENMESİ

Öz

Bu araştırmada eğitim bilimleri alanında yapılan ölçek geliştirme çalışmalarının uluslararası standartlar kapsamında incelenmesi amaçlanmıştır. Araştırmada nitel araştırma yaklaşımlarından döküman incelemesi yöntemi kullanılmıştır. Araştırma sürecinde eğitim bilimleri alanında ve Türkiye kökenli ölçek geliştirme çalışmaları incelenmiştir. Araştırma kapsamına 2010-2023 yılları arasındaki toplam 85 çalışma dâhil edilmiştir. Araştırma kapsamında konu alanı, örneklem büyüklüğü, örneklem türü, geçerlik ve güvenilirlik analizleri, madde ve faktör eleme kriterleri, kullanılan istatistik yazılımları ve araştırmacı statüsü gibi kriterler dikkate alınmıştır. Araştırma sonucunda ölçek geliştirme çalışmalarının büyük bir çoğunluğunun uluslararası standartları karşılamadığı ya da belirli düzeyde standartları karşılayabildiği belirlenmiştir. Örneklem türü değişkeni açısından oldukça farklı türlerde çalışmalar yapıldığı görülmektedir. Örneklem büyüklüğü sonuçları incelendiğinde örneklemelerin yetersiz olduğu belirlenmiştir. Geçerlik ve güvenilirlik analizlerinin büyük oranda kullanıldığı ancak yetersiz olduğu tespit edilmiştir. Madde ve faktör eleme konusunda kısıtlı oranda bilgi verildiği ya da hiç yer verilmediği anlaşılmıştır. Kullanılan istatistik yazılımları konusunda tüm çalışmalarda yer verildiği görülürken, araştırmacı statüsü gibi konularda pek fazla yer verilmediği görülmektedir. Araştırma sonucunda çeşitli önerilerde bulunulmuştur.

Anahtar kelimeler: Eğitim bilimleri; ölçek geliştirme; uluslararası standartlar

GİRİŞ

Günümüzde eğitim ve öğretim hizmetleri sunulurken daha kaliteli bir şekilde hizmet verilebilmesi adına birçok çalışma yapılmaktadır. Bu çalışmalar bireysel, örgütsel, toplumsal ve istatistiksel açılardan çeşitlilik göstermektedir. Eğitim hizmetlerinin yoğun olarak sunulduğu alanlardan birisi de yükseköğretim alanıdır. Yükseköğretimde topluma yön verebilecek, bilimsel temelleri bulunan ve akademik açıdan yetkin araştırmacıların gerçekleştirdiği çalışmalar bulunmaktadır. Bu çalışmaların zaman zaman kontrol edilmesi, farklı açılardan ele alınması ve niteliklerinin incelenmesi hem daha kaliteli bir eğitim hizmetinin sunulmasına yardımcı olmakta hem de bilimsel gelişmelerin sistematik olarak iyileştirilmesine katkı sunmaktadır (Mor-Dirlik, 2014). Belirli alan ve eğitim kademelerinde elde edilen bilimsel bulgular ve bilimsel bilgi birikimi ışığında ölçme alanında geliştirilen yeni yöntem, araç ve teknikler diğer alan ve eğitim kademelerinde de yaygınlaşmıştır (Uçan, 2005). İlkokuldan başlayarak yükseköğretime kadar hemen hemen her düzeyde öğrencilerin bilişsel, duyuşsal ve psikolojik özellikleri ölçülmektedir (Çüm, 2018; Yılmaz, 2021). Bu ölçüm işlemleri yapılırken genellikle öğrenciler bir test ile karşılaştırılmakta ve öğrencilerin testlere verdikleri tepkiler belirli bir yöntem kullanılarak puanlanmaktadır (Anastasi, 1988). Bu testler içerisinde psikolojik testler, ölçekler ve anketler sıklıkla kullanılmaktadır. Çünkü bu testler objektif sonuçlar sunabilmekte, geçerliği ve güvenilirliği yüksek sonuçlar verebilmekte ve kolaylıkla uygulanabilmektedir (Conway, 2006).

Bireylerin sahip olduğu özelliklerin neler olduğunun belirlenebilmesi amacıyla öncelikle ölçümler yapılmalıdır. Bu ölçümlerde alan yazında geçerliği ve güvenilirliği sağlanmış araçlarla yapılmalıdır. Yükseköğretimde psikolojik testler yoğun olarak kullanılmakla birlikte bunlar içerisinde ölçekler en çok tercih edilen veri toplama araçları olarak görülmektedir. Araştırmacılar çalışmalarında çok farklı türlerde ölçeklere ihtiyaç duyabilmektedir. Bu konuda ya alan yazında bulunan hazır ölçekleri kullanmakta ya da ihtiyaçlarına göre ölçek geliştirmekte veyahut uyarlamaktadırlar. Ancak bu durum her zaman sağlıklı bir şekilde gerçekleştirilmemektedir. Çünkü ölçekler sistematik ve belirli bir standartları karşıladığı durumlarda amacına hizmet edebilmektedir (Özoğlu, 2007). Bir başka ifadeyle elde edilen bilgilerin anlamlı ve kullanılabilir olabilmesi için bu ölçeklerin titizlikle, standartlara uygun ve belirli aşamalardan geçerek oluşturulması gerekmektedir (Baykul & Turgut, 2010). Eğitim bilimleri alanında son yıllarda ölçek geliştirme çalışmalarında yoğun bir artış olduğu görülmektedir. Bu durum gerek bilimsel araştırma ihtiyacının karşılanması gerekse de bu alanda yeni ve özgün araştırmalara imza atılması amacıyla sıklıkla tercih edilmektedir. Ölçek geliştirmeyi tercih eden araştırmacılar ölçmek istedikleri özelliğin içeriğini ve yapısını çok iyi bir şekilde tanımalıdır. Bunu yaparken de daha önce yapılmış ölçek geliştirme çalışmalarını inceleyerek belirli adımlarla bu işlemi gerçekleştirmelidir. Ancak ölçek geliştirme çalışmaları uzun soluklu ve deneyim gerektiren araştırmalardır. Aynı zamanda ölçek geliştirmeye niyetli olan kişilerin bu alanda daha önce çalışma yapmış olması, alan yazına hâkim olması, yeterli düzeyde istatistiki bilgiye sahip olması gerekmektedir (Gül & Sözbilir, 2015). Bunlara ek olarak geliştirilen ölçeklerin yeni gelişmelere bağlı olarak yeniden revize edilmesi, farklı örneklerde kullanılması ve psikometrik ölçümlerinin kontrol edilmesi ve bunların yenilenmesi oldukça önemlidir (Azaltun, 2008). Alan yazında birçok ölçek geliştirme çalışması ve bu çalışmalara yön verecek nitelikte olan çalışmalarda yer almaktadır (Ayyıldız & Yılmaz, 2021; Cabrera-Nyugen,

2010; Carpenter, 2018; DeVellis & Thorpe, 2021; DiStefano, Zhu ve Mindrila, 2009; Erkuş, 2012; Knafl et al., 2012; Simms, 2008). Buna rağmen araştırmacılar üzerinde yayın baskısının olması ve alanda ölçek geliştirme çalışmalarının kolay bir çıkış kapısı olarak görülmesi bu araştırmaların artmasına ve beraberinde birçok hatayı ve noksanlıkları da getirmektedir (Erkuş, 2007). Araştırmacılar zaman zaman ölçek geliştirmek yerine ölçek uyarlamayı da tercih edebilmektedir (Yılmaz, 2018). Çünkü ölçek uyarlama ölçek geliştirmekten hem daha kolay hem de daha hızlı bir şekilde gerçekleştirilebilmektedir (Hambleton & Patsula, 1999). Eğitim bilimleri alanında birbirinden bağımsız nitelikte ve farklı ölçek geliştirme süreçlerinin kullanıldığı birçok çalışma bulunmaktadır. Dolayısıyla bu farklılıklar zamanla kalıplaşmakta ve yazılı olmayan kurallar şeklinde kabul edilebilmektedir. Ancak ölçek geliştirme/uyarlama çalışmalarında sistematik ve kabul edilebilir bir yöntem izlenilmesi gerekmektedir (Acar-Güvendir & Özer-Özkan, 2015). Her geçen gün yeni bir ölçeğin oluşturulması ve meydana gelen bu bilgi yığınının anlamlı bir şekilde yorumlanabilmesi için güvenilir, geçerli ve kapsamlı sayılabilecek üst çalışmalara ihtiyaç duyulmaktadır (Akgöz, Ercan & Kan, 2004).

Alan yazın incelendiğinde zaman zaman üst çalışmaların yapıldığı görülmektedir. Çüm (2013) yapmış olduğu çalışmada eğitim ve psikoloji alanında yapılan makale çalışmalarını incelemiştir. Mor-Dirlik (2014) eğitim bilimleri kapsamında hazırlanan beş farklı tezi farklı açılardan gözlemlemiştir. Gül ve Sözbilir (2015) fen ve matematik alanındaki ölçek geliştirme çalışmalarını tematik olarak ele almıştır. Tosun ve Taşkesenligil (2014) fen eğitimi kapsamında geliştirilen çalışmaları incelemiştir. Boztunç-Öztürk, Eroğlu ve Kelecioğlu (2015) eğitim alanında gerçekleştirilen ölçek uyarlama makalelerini irdelemiştir. Güler ve Ayan (2020) 2002-2018 yılları arasında Türkiye'de eğitim bilimleri alanında yapılan 112 tutum ölçeği geliştirme çalışmasını incelemiş ve tutum ölçeklerinin geliştirilmesinde dikkate alınması gereken işlem basamaklarının ne ölçüde karşılandığını araştırmıştır. Görüldüğü üzere ölçek geliştirme alanında üst çalışmaların yapıldığı ve bu çalışmaların sık sık tekrar edilmesi gerekmektedir. Bu nedenle eğitim bilimleri alanında yapılan ölçek geliştirme çalışmalarının yeniden incelenmesinin alan yazına katkısı bulunmaktadır. Bu araştırmanın temel amacı eğitim bilimleri alanında yapılan ölçek geliştirme çalışmalarının uluslararası standartlar kapsamında incelenmesidir. Bu kapsamda şu alt problemlere yanıt aranmıştır:

1. Ölçek geliştirme konusunda hangi konu alanları tercih edilmiştir?
2. Örneklem türü ve örneklem büyüklüğü konusunda ölçek geliştirme çalışmaları ne gibi farklılıklar oluşturmaktadır?
3. Ölçek geliştirme çalışmalarında ne tür geçerlik analizleri kullanılmıştır?
4. Ölçek geliştirme çalışmalarında ne tür güvenirlik analizleri kullanılmıştır?
5. Ölçek geliştirme sürecinde kullanılan madde eleme kriterleri nelerdir?
6. Ölçek geliştirme sürecinde kullanılan faktör eleme kriterleri nelerdir?
7. Ölçek geliştirme aşamasında ne tür istatistik yazılımları kullanılmıştır?
8. Ölçek geliştirme çalışmalarında araştırmacı statüsü hakkında ne düzeyde bilgi verilmiştir?

YÖNTEM

Bu arařtırmada Türkiye’de 2010-2023 yılları arasında yapılan ölçek geliřtirme çalıřmaları ierik analizi yöntemi yardımıyla incelenmiřtir. İerik analizi çalıřmaları kavramların ve olguların dokümanlar yardımıyla incelenmesinde sıklıkla kullanılmaktadır (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz ve Demirel, 2008). İerik analizi genel olarak üç farklı başlık altında toplanabilir. Bu başlıklar, meta-analiz, meta sentez ve betimsel ierik analizi şeklinde ifade edilebilir (Çalık ve Sözbilir, 2014).

Bu arařtırma ile eğitim bilimleri alanında yapılan ölçek geliřtirme çalıřmaları belirli deęişkenler açısından incelenmiş ve uluslararası standartlar kapsamında karşılaştırılmıştır. Bu özellięi ile arařtırma betimsel ierik analizi olarak ifadelendirilebilir.

İerik Analizine Tabi Tutulan Ölçek Geliřtirme Çalıřmaları

İerik analizine tabi tutulan ölçek geliřtirme çalıřmaları 2010-2023 yılları arasında yapılmıřtır. Çalıřmalar Türkiye özelinde yapılmıř olan ölçek geliřtirme çalıřmalarıdır. Arařtırmada yalnızca eğitim bilimleri alanına giren ölçek geliřtirme çalıřmaları dikkate alınmıřtır.

Verilerin Toplanması

Arařtırma sürecinde elde edilen ölçek geliřtirme çalıřmaları 2010-2023 yılları arasında Türkiye özelinde gerekleřtirilen çalıřmalardan meydana gelmektedir. Söz konusu çalıřmalar aranırken “eęitim bilimleri, ölçek geliřtirme, geçerlik ve güvenilirlik çalıřması” gibi anahtar kelimeler kullanılmıřtır. Aynı zamanda TRDizin-ULAKBİM, ERIC, H.W.Wilson gibi alan indeksleri ve SSCI düzeyi indekslerde bulunan yayınlarda incelenmiřtir. Bu veritabanlarının incelenmesinin nedeni, bu alanlarda yayınlanan çalıřmaların belirli bir standarda sahip olması ve belirli ölçütlere göre deęerlendirmeye tabi tutulmasıdır. Arařtırma sürecinde ölçek geliřtirme makalelerinin incelendięi dergilerde bazı kriterler de dikkate alınmıřtır. Bu kriterler řunlardır:

- Dergilerin impact faktörlerinin yüksek olması,
- Seçilen dergilerin TUBİTAK yayın teřvik kriterlerine baęlı olarak yüksek nitelikte bulunması,
- Uluslararası geçerlilięi bulunan indekslere sahip olması,
- Dergilerin eğitim bilimleri alanında yayın yapıyor olması,
- 2010-2023 yılları arasında düzenli olarak dergilerde yayın yapıyor olması,
- Dergi editörleri ve alan editörlerinin ölçme ve deęerlendirme alanlarında yetkin olması,
- Dergilerin köklü ve uzun bir yayın gemişine sahip olması.

Verilerin Analizi

Arařtırma kapsamında elde edilen veriler ierik analizine tabi tutulmuş olup kategorisel sınıflandırma ve frekans analizi sunulmuřtur. Bu doęrultuda uygulama verileri betimsel olarak analiz edilmiş, frekans ve yüzdelik oranlar

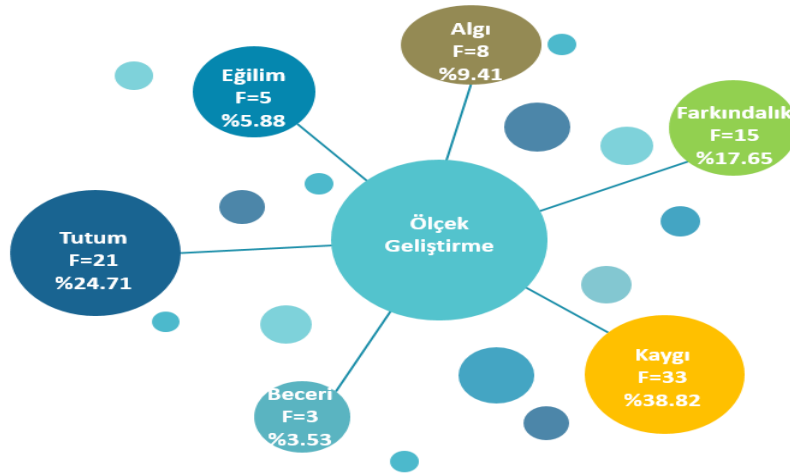
yardımla görselleştirilmiştir. Ayrıca araştırmacılar tarafından çalışmalarda elde edilen sonuçlar kodlamalar yapılarak kategorilere ayrılmıştır.

Etik Kurallara Uygunluk ve Gerekli İzinlerin Alınması

Araştırma sürecinin tamamında etik kurallara azami düzeyde dikkat edilmiş olup gerekli etik kurul izinleri alınmıştır. Araştırmanın etik kurul izni Kastamonu Üniversitesi Sosyal ve Beşeri Bilimler Araştırma ve Yayın ve Etik Kurulu tarafından 04.05.2020 tarih ve 30 sayılı karar ile onaylanmıştır.

BULGULAR

Araştırma sonuçları her bir alt problem durumu için ayrı ayrı incelenmiştir. Araştırmanın ilk problem durumu “ölçek geliştirme konusunda hangi konu alanları tercih edilmiştir?” şeklindedir. Şekil 1’de tercih edilen konu alanları bulunmaktadır.



Şekil 1. Ölçek Geliştirme Konu Alanları

Şekil 1 incelendiğinde ölçek geliştirme çalışmalarının konu alanlarına göre kaygı, tutum, farkındalık, algı, eğilim ve beceri şeklinde ayrıldığı görülmektedir. Konu alanlarına göre en çok geliştirilen ölçek türü kaygı ölçeği iken en az geliştirilen ölçek türü ise beceri olarak belirlenmiştir. Araştırmanın ikinci problem durumu “örneklem türü ve örneklem büyüklüğü konusunda ölçek geliştirme çalışmaları ne gibi farklılıklar oluşturmaktadır?” şeklindedir. Tablo 1’de örneklem türü ve örneklem büyüklüğüne yönelik bulgular sunulmuştur.

Tablo 1. Örneklem Türü ve Örneklem Büyüklüğüne Yönelik Bulgular

Örneklem Türü	(f)	%
Okul öncesi düzeyi	3	3.53
İlköğretim düzeyi	18	21.18
Ortaöğretim düzeyi	28	32.94
Lisans düzeyi	22	25.88
Lisansüstü düzeyi	14	16.47
Toplam	85	100

Örneklem Büyüklüğü	(f)	%
0-100 arası	2	2.35
101-150 arası	7	8.24
151-250 arası	11	12.94
251-400 arası	16	18.82
401-600 arası	17	20.00
601-800 arası	13	15.29
801-1000 arası	11	12.94
1000 ve üzeri	8	9.41
Toplam	85	100

Tablo 1 incelendiğinde ölçek geliştirme çalışmalarının örneklem türüne göre okul öncesi, ilköğretim, ortaöğretim, lisans ve lisansüstü düzeylerde tercih edildiği belirlenmiştir. Örneklem türleri içerisinde en çok ortaöğretim düzeyi en az ise okul öncesi düzeyi tercih edilmiştir. Ölçek geliştirme çalışmaları örneklem büyüklüğüne göre incelendiğinde birçok farklı büyüklükte örneklem seçildiği görülmektedir. Örneklem büyüklükleri içerisinde en çok tercih edilen büyüklük 401-600 arası olurken en az tercih edilen büyüklük ise 0-100 arası olmuştur. Araştırmanın üçüncü problem durumu “ölçek geliştirme çalışmalarında ne tür geçerlik analizleri kullanılmıştır?” şeklindedir. Tablo 2’de geçerlik analizlerine yönelik bulgular sunulmuştur.

Tablo 2. Geçerlik Analizine Yönelik Bulgular

Geçerlik Analizi	(f)	%
Kapsam geçerliği	15	17.65
Görünüş geçerliği	11	12.94
Ölçüt geçerliği	22	25.88
Yapı geçerliği	37	43.53
Toplam	85	100

Tablo 2 incelendiğinde ölçek geliştirme çalışmalarında geçerlik analizine yönelik olarak kapsam, görünüş, ölçüt ve yapı geçerliliklerinin analiz edildiği görülmektedir. Bu analizler içerisinde en çok yapı geçerliği incelenirken en az ise görünüş geçerliği incelenmiştir. Ölçüt geçerliği ve kapsam geçerliği de tercih edilen diğer geçerlik analizi türleridir. Araştırmanın dördüncü problem durumu “ölçek geliştirme çalışmalarında ne tür güvenilirlik analizleri kullanılmıştır?” şeklindedir. Tablo 3’te güvenilirlik analizlerine yönelik bulgular sunulmuştur.

Tablo 3. Güvenirlik Analizine Yönelik Bulgular

Güvenirlik Analizi	(f)	%
Test-tekrar test	22	25.88
Eşdeğer formlar	11	12.94
İç tutarlılık	36	42.35
Testi yarılama	12	14.12
Alt-Üst grup analizi	4	4.71
Toplam	85	100

Tablo 3 incelendiğinde ölçek geliştirme çalışmalarında güvenilirlik analizine yönelik olarak test-tekrar test, eşdeğer formlar, iç tutarlılık, testi yarılama ve alt-üst grup analizlerinin yapıldığı görülmektedir. Bu analizler

içerisinde en sık tercih edilen iç tutarlılık en az tercih edilen ise alt-üst grup analizi olmuştur. İç tutarlılık analizleri kendi içerisinde farklı alt türlere ayrılmakla birlikte en sık tercih edilen analiz türüdür.

Araştırmanın beşinci problem durumu “ölçek geliştirme sürecinde kullanılan madde eleme kriterleri nelerdir?” şeklindedir. Tablo 4’te madde eleme kriterlerine yönelik bulgular sunulmuştur.

Tablo 4. Madde Eleme Kriterlerine Yönelik Bulgular

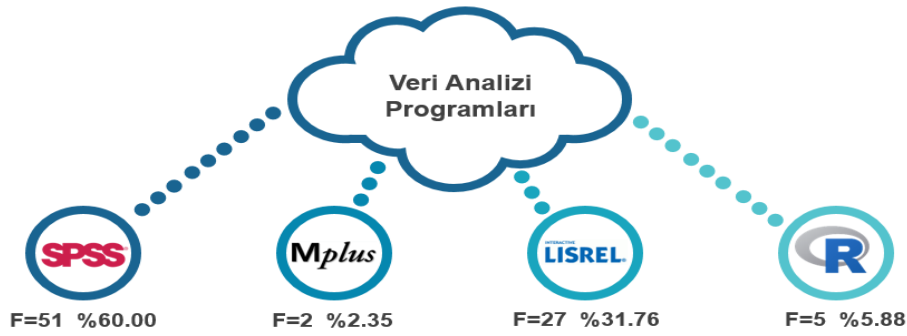
Madde Eleme Kriterleri	(f)	%
Madde faktör yükleri	29	34.12
Binişik yükler	14	16.47
Madde analizi	22	25.88
Uzman görüşü	7	8.24
Madde-toplam korelasyonu	8	9.41
Ortak varyanslar	5	5.88
Toplam	85	100

Tablo 4 incelendiğinde ölçek geliştirme çalışmalarında madde eleme kriterlerine yönelik olarak madde faktör yükleri, binişik yükler, madde analizi, uzman görüşü, madde-toplam korelasyonu ve ortak varyansların kullanıldığı görülmektedir. Bu kriterler içerisinde en çok tercih edilen madde faktör yükleri, en az tercih edilen ise ortak varyanslardır. Araştırmanın altıncı problem durumu “ölçek geliştirme sürecinde kullanılan faktör eleme kriterleri nelerdir?” şeklindedir. Tablo 5’te faktör eleme kriterlerine yönelik bulgular sunulmuştur.

Tablo 5. Faktör Eleme Kriterlerine Yönelik Bulgular

Faktör Eleme Kriterleri	(f)	%
Özdeğer sonuçları	41	48.24
Yamaç birikinti grafiği	23	27.06
Varyans oranı	15	17.65
Madde sayısı	2	2.35
Uzman görüşü	2	2.35
Alan yazın sonuçları	2	2.35
Toplam	85	100

Tablo 5 incelendiğinde ölçek geliştirme çalışmalarında faktör eleme kriterlerine yönelik olarak özdeğer sonuçları, yamaç birikinti grafiği, varyans oranı, madde sayısı, uzman görüşü ve alan yazın sonuçlarının tercih edildiği görülmektedir. Bu kriterler içerisinde en çok tercih edilen madde özdeğer sonuçları olmuştur. En az tercih edilen kriterler ise madde sayısı, uzman görüşü ve alan yazın sonuçlarıdır. Araştırmanın yedinci problem durumu “ölçek geliştirme aşamasında ne tür istatistik yazılımları kullanılmıştır?” şeklindedir. Şekil 2’de ölçek geliştirme sürecinde kullanılan istatistik yazılımlarına yönelik bulgular sunulmuştur.



Şekil 2. Ölçek Geliştirme Sürecinde Kullanılan İstatistik Yazılımları

Şekil 2 incelendiğinde ölçek geliştirme sürecinde kullanılan istatistik yazılımlarına göre SPSS, Mplus, LISREL ve R programlarının tercih edildiği görülmektedir. Bu yazılımlar içerisinde SPSS programı en çok tercih edilen program olmuştur. Daha sonra bu durumu LISREL, R ve Mplus programları takip etmiştir. Araştırmancının sekizinci problem durumu “ölçek geliştirme çalışmalarında araştırmacı statüsü hakkında ne düzeyde bilgi verilmiştir?” şeklindedir. Tablo 6’da araştırmacı statüsüne yönelik bulgular sunulmuştur.

Tablo 6. Araştırmacı statüsüne yönelik bulgular

Araştırmacı Statüsü	(f)	%
Yüksek lisans öğrencisi	23	27.06
Doktora öğrencisi	26	30.59
Akademisyen	34	40.00
Belirtilmemiş	2	2.35
Toplam	85	100

Tablo 6 incelendiğinde ölçek geliştirme çalışmalarında araştırmacıların yoğun olarak akademisyenlerden ve lisansüstü düzeyde eğitim gören öğrencilerden meydana geldiği belirlenmiştir. 2 çalışmada ise araştırmacı statüsü belirtilmemiştir.

TARTIŞMA, SONUÇ ve ÖNERİLER

Eğitim bilimleri alanında yapılan ölçek geliştirme çalışmalarının uluslararası standartlar çerçevesinde değerlendirmesinin yapıldığı bu çalışmada, ilk olarak konu alanları incelenmiştir. 85 adet ölçek geliştirme çalışması incelendiğinde çalışmaların yoğun olarak kaygı, tutum ve farkındalık konularında hazırlandığı görülmektedir. Bu durum alan yazında da birçok çalışmada vurgulanmıştır (Gül & Sözbilir, 2015; Tosun & Taşkesenligil, 2014). Çünkü kaygı, tutum ve farkındalık bireylerin davranışlarını en kolay yansıtabildiği ve kritik sonuçlar elde edilebilen davranışlardır. Benzer şekilde Çiltaş, Güler ve Sözbilir (2012) tarafından yapılan bir diğer çalışmada da tutum ve kaygı çalışmalarının ölçek geliştirme çalışmalarında sıklıkla tercih edildiği vurgulanmaktadır. Ölçek geliştirme çalışmalarında en az karşılaşılan ölçek türü beceri ölçekleridir. Çünkü beceriler genellikle uzun soluklu gözlemler ve belirli bir zaman diliminde yapılan düzenli ölçümlerle gözlemlenebilmektedir. Bu nedenle beceri ölçeklerinin az olması olağan bir durum olarak karşılanabilir.

Ölçek geliştirme çalışmalarında ikinci değişken örneklem türü ve örneklem büyüklüğüdür. Örneklem türü incelendiğinde okul öncesi dönemden başlayarak yükseköğretime kadar birçok farklı düzeyde çalışmalar yapıldığı görülmektedir. Bu çalışmalar içerisinde okul öncesi düzeyde bulunan çalışmalar oldukça azdır. Çünkü okul öncesi dönemde bulunan çocukların ölçek sorularına yanıt vermesi çoğu zaman zor bir davranıştır. Aynı zamanda uzun süre dikkatlerini toplayamamaları da bir diğer faktördür. Örneklem türü olarak en çok çalışma ortaöğretim düzeyinde gerçekleştirilmiştir. Bunun nedeni lise düzeyinin bir geçiş aşaması olması ve öğrencilerin yükseköğretime geçmeden önce hem mesleki kaygılarının olması hem de gelecek beklentilerinin yüksek olmasıdır (Gençtürk & Memiş, 2010). Bu sonuçlar Tosun ve Taşkesenligil (2014) tarafından yapılan çalışma sonuçları ile desteklenmektedir. Örneklem türüne göre en çok çalışma yapılan diğer grup ise lisans düzeyinde bulunan öğrencilerdir. Bu durumun temel nedeni lisans düzeyinde bulunan öğrencilerin kolay ulaşılabilir bir örneklem grubu olmasıdır. Aynı zamanda kısa sürede çok fazla sayıda kişiye ulaşılabilmesi de bu grubun diğer önemli özelliğini oluşturmaktadır (Deveci, Çalmaz & Açık, 2012). Ölçek geliştirme konusunda konu alanı ve örneklem türü değişkenlerinde hali hazırda bir uluslararası standart bulunmamaktadır. Ancak psikolojik test ölçümleri açısından konuya yaklaşıldığında konu alanı ve örneklem türünün çeşitlilik ve zenginlik arz etmesi alan yazını güçlendiren bir özellik olarak değerlendirilebilir.

Ölçek geliştirme çalışmaları örneklem büyüklüğü açısından incelendiğinde en yoğun örneklem grubunun 401-600 kişi arasında olduğu görülmektedir. 1000 ve üzeri kişi sayısına ulaşılan çalışmanın ise düşük seviyelerde olduğu görülmektedir. Örneklem büyüklüğü konusunda alan yazında bazı standartlar bulunmaktadır. Bu standartlar ilk olarak madde sayısı ile ilgilidir. Uluslararası ölçek geliştirme standartları dikkate alındığında örneklem büyüklüğünün soru sayısının en az 5 katı ile 10 katı arasında olması gerektiği ifade edilmektedir (Özdamar, 2003). Bu durum özellikle kullanılan istatistik programlarının örneklem büyüklüğüne hassas olması nedeniyle çoğu zaman yüksek tutulmaya çalışılmaktadır. Ancak örneklem büyüklüğü bir noktadan sonra sonuçlarda değişiklik göstermemektedir. Buna ek olarak örneklem büyüklüğü arttıkça ilişkisi olmayan sorular ilişki göstermeye başlamakta ve bu da ölçeğin amacının dışına çıkılmasına neden olabilmektedir (Yılmaz, 2018). Ölçek geliştirme sürecinde örneklem büyüklüğüne yönelik bir diğer kural ise Comrey ve Lee (1992) tarafından belirlenmiştir. Buna göre örneklem sayısı 100 kişi ise zayıf, 200 kişi orta, 300 kişi iyi, 500 kişi çok iyi ve 1000 kişi mükemmel olarak ifade edilmiştir. Ancak burada unutulmaması gereken nokta ölçeğin sahip olduğu madde sayısı ve veri doygunluğu sınırlandırıcıdır. Veri doygunluğuna ulaşıldıktan sonra örneklem sayısını arttırmak pek fazla sonuçları değiştirmeyecektir (Delice, 2010). Araştırma kapsamında incelenen çalışmalar dikkate alındığında birçok çalışmanın örneklem büyüklüğünün soru sayısına göre yetersiz olduğu ve ölçeklerde kabul edilen değer aralıklarının yeterince iyi olmadığı belirlenmiştir.

Araştırma sonuçları geçerlik analizleri kapsamında incelendiğinde en çok tercih edilen geçerlik türünün yapı geçerliği olduğu görülmektedir. Özellikle 2000'li yıllardan sonra yapılan ölçek geliştirme çalışmalarında bilgisayar teknolojilerinin de yardımıyla birçok çalışmada açımlayıcı ve doğrulayıcı faktör analizleri yapılmaya başlanmıştır (Erkuş, 2012; Yurdugül, 2005). Alan yazında birçok çalışmada doğrulayıcı faktör analizi yapılmasının gerekliliği belirtildiğinden yapı geçerliğine yönelik kontrollerin yoğun olması doğal bir sonuç olarak karşılanabilir

(Brinkmann, 2009). Bu durumu ölçüt geçerliği ve kapsam geçerliği takip etmektedir. Nitekim alan yazında bulunan çalışmalarda kapsam geçerliliğinin yapılması gerektiği birçok kez ifade edilmiştir (Slavec & Drnovsek, 2012). Geçerlik analizleri uluslararası standartlar dikkate alınarak incelendiğinde kapsam, görünüş ve yapı geçerliğinin yapılmasının gerekli olduğu ifade edilmektedir (Yurdugül & Bayrak, 2012).

Güvenirlik analizi kapsamında elde edilen sonuçlar incelendiğinde çalışmaların yarısından azında iç tutarlık katsayısının hesaplandığı görülmektedir. Bunun dışında test tekrar test yöntemi ve split-test yöntemi sıklıkla kullanılmaktadır. Alan yazında öne çıkan ve belirli standartları sağlamış olan dergilerde iç tutarlık katsayısının hesaplanması çoğu zaman zorunlu olarak istenilmektedir (Topu, Baydaş, Turan & Göktaş, 2013). Öğretim teknolojisi araştırmalarında geçerlik ve güvenilirlik önlemleri. Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 42(1), 110-126. Bu nedenle incelenen çalışmaların güvenilirlik analizi açısından yeterli kriterleri sağlamadığı ifade edilebilir.

Araştırma kapsamında madde eleme kriterleri incelendiğinde en sık tercih edilen yöntemin madde faktör yüklerinin hesaplanması ve madde analizi seçeneği olduğu görülmektedir. Worthington ve Whittaker (2006) tarafından yapılan çalışmada madde faktör yüklerinin sıklıkla kullanıldığı ifade edilmektedir. Aynı zamanda uluslararası standartlar kapsamında dikkate alınması gereken bir diğer faktör ise madde analizidir. Madde analizi ile ölçekler alt boyutlara daha iyi bir şekilde ayrılabilir (Tavşancıl, 2002). İncelenen çalışmalar kapsamında madde faktör yükü kullanımı ve madde analizi kullanımının oldukça az olduğu görülmektedir. Araştırmada faktör eleme kriterleri incelendiğinde en sık kullanılan yöntemi özdeğer sonuçları olduğu görülmektedir. Bu durumu scree plot grafiği takip etmektedir. Alan yazında özdeğer sonuçlarının ve scree plot grafiklerinin incelenmesi sıklıkla vurgulanmaktadır. Hinkin (1998) tarafından yapılan çalışmada uluslararası standartlar kapsamında bu değerlerin incelenmesi gerektiği vurgulanmaktadır. Aynı zamanda Erkuş (2012) yapmış olduğu çalışmada özdeğer sonuçlarının tek başına yeterli olmadığını aynı zamanda scree plot grafiği sonuçlarının da değerlendirmeye alınması gerektiğini ifade etmektedir. Bu sonuçlar ışığında incelenen çalışmaların özdeğer sonuçları açısından kısmen yeterli olduğu ifade edilebilir.

Ölçek geliştirme sürecinde yoğun istatistiksel işlemler gerçekleştirilmektedir. Bu konuda alan yazında birçok yazılım bulunmaktadır. Ancak en çok tercih edilen yazılımlar arasında SPSS ve LISREL programları yer almaktadır. Bu programlara ek olarak Mplus ve R gibi yazılımlarda tercih edilmektedir. Araştırma sonucunda en sık tercih edilen yazılım SPSS olmuştur. Bunun nedeni SPSS programının kullanıcı arayüzünün kolay ve kullanışlı olması gerekçe olarak gösterilebilir. Ancak LISREL, Mplus ve R programlarında ise kullanıcı arayüzü biraz daha fazla bilgisayar okuryazarlığı becerileri ve kodlama becerileri gerektirmektedir. Bu nedenle SPSS programının yoğun olarak tercih edilmesi beklenen bir sonuçtur. Araştırma sürecinin son değişkeni araştırmacı statüsüdür. Ölçek geliştirme çalışmaları çoğu zaman yoğun emek, deneyim ve yeterli alan bilgisine sahip olmayı istemektedir. Araştırma sonuçları incelendiğinde çalışmaların büyük bir çoğunluğunun araştırmacılar, yüksek lisans ve doktora öğrencileri tarafından geliştirildiği görülmektedir. Bu noktada yüksek lisans ve doktora öğrencilerinin sayısının çok olması kısmen de olsa endişe yaratabilmektedir. Çünkü henüz tam olarak yetişmemiş olan

araştırmacıların ölçek geliştirmesi alana katkı sağlayabileceği gibi hatalı ölçeklerin üretilmesine de neden olabilecektir.

Tüm bu bilgiler ışığında şu önerilere yer verilebilir:

1. Ölçek geliştirme çalışmalarında yeni bir ölçek geliştirilmeden önce alan yazının iyi bir şekilde taranması ve amaca uygun olan ölçeklerin kullanılmasına özen gösterilmelidir. Mecbur olunmadığı sürece yeni bir ölçek geliştirme sürecine girilmemesi gerekmektedir.
2. Örneklem türü konusunda alan yazında yetersizliği bulunan alanlarda ölçek geliştirme çalışmaları yapılmalı ve zenginlik oluşturulmalıdır.
3. Ölçek geliştirme çalışmalarında örneklem büyüklüğü önemli bir standarttır. Bu konuda yeterli düzeyde örneklem büyüklüğüne ulaşılması için gerekli istatistiksel teknikler kullanılmalıdır.
4. Araştırma sürecinde kapsam ve yapı geçerliği olmazsa olmaz kriterler arasında yer almaktadır. Bu geçerlik türlerinin kullanılmasına özen gösterilmelidir.
5. Ölçek geliştirme süreci deneyimli ve alan bilgisine sahip kişiler tarafından yapılmalıdır. Aksi takdirde birbirine benzer düzeyde birçok ölçek geliştirilmekte ve alan yazında aynı amaca hizmet eden birçok ölçek yığını oluşmaktadır.

Etik Metni

“Bu makalede dergi yazım kurallarına, yayın ilkelerine, araştırma ve yayın etiği kurallarına, dergi etik kurallarına uyulmuştur. Makale ile ilgili doğabilecek her türlü ihlallerde sorumluluk yazar(lar)a aittir. Makalenin etik kurul izni Kastamonu Üniversitesi/Yayın Etiği Kurulu tarafınca 04.05.2020 tarih ve 30 sayılı kararı ile alınmıştır.”

Yazar(lar)ın Katkı Oranı Beyanı: Bu çalışmada birinci yazarın katkı oranı %50, ikinci yazarın katkı oranı %50'dir.

KAYNAKÇA

- Acar-Güvendir, M., & 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.
- Akgöz, S., Ercan, İ., & Kan, İ. (2004). Meta-analizi. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 30(2), 107-112.
- Anastasi, A. (1988). *Psychological testing*. (6th Ed.). MacMillan Publishing Co. Inc.
- Ayyıldız, P., & Yılmaz, A. (2021). 'Moving the kaleidoscope' to see the effect of creative personality traits on creative thinking dispositions of pre-service teachers: The mediating effect of creative learning environments and teachers' creativity fostering behavior. *Thinking Skills and Creativity*, 41, 100879, 1-10. <https://doi.org/10.1016/j.tsc.2021.100879>
- Azaltun, M. (2008). VI. araştırma yöntemleri semineri-ölçme ve ölçek geliştirme. *Anatolia: Turizm Araştırmaları Dergisi*, 19(1), 104-111.
- Baykul, Y., & Turgut, F. (2010). *Eğitimde ölçme ve değerlendirme*. Pegem Akademi Yayıncılık.

- Boztunç-Öztürk, N., Eroğlu, M.G., & Kelecioğlu, H. (2015). Eğitim alanında yapılan ölçek uyarlama makalelerinin incelenmesi. *Eğitim ve Bilim*, 40(178), 123-137. <https://doi.org/10.15390/EB.2015.4091>
- Brinkman, W. P. (2009). Design of a questionnaire instrument. S. Love (Edt). *Handbook of mobile technology research methods* içinde (ss.31-53). Nova Publisher.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2008). *Bilimsel araştırma yöntemleri*. Pegem Akademi Yayınları.
- 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.
- Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, 12(1), 25-44. <https://doi.org/10.1080/19312458.2017.1396583>
- Comrey, A. L., & Lee, H. L. (1992). *A first course in factor analysis*, Hillsdale, Erlbaum.
- Conway, M. J. (2006). *How to collect data: measurement and evaluation*. American Society for Training and Development.
- Çalık, M., & Sözbilir, M. (2014). İçerik analizinin parametreleri. *Eğitim ve Bilim*, 39(174), 33-38. <https://doi.org/10.15390/EB.2014.3412>
- Çiltaş, A., Güler, G., & Sözbilir, M. (2012). Türkiye'de matematik eğitimi araştırmaları: İçerik analizi çalışması. *Kuram ve Uygulamada Eğitim Bilimleri*, 12(1), 515-580.
- Çüm. S. (2013). *Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi*. [Yayımlanmamış Yüksek Lisans Tezi]. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Delice, A. (2010). Nicel araştırmalarda örneklem sorunu. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(4), 1969-2018.
- Deveci, S. E., Çalmaz, A., & Açıık, Y. (2012). Doğu Anadolu'da yeni açılan bir üniversitenin öğrencilerinde kaygı düzeylerinin sağlık, sosyal ve demografik faktörler ile ilişkisi. *Dicle Tıp Dergisi*, 39(2), 189-196.
- DeVellis, R.F., & Thorpe, C.T. (2021). *Scale development: Theory and applications*. Sage publications.
- 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.
- Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar. *Türk Psikoloji Bülteni*, 13(40), 17-25.
- Erkuş, A. (2012). *Psikolojide ölçek ve ölçek geliştirme-I: Temel kavramlar ve işlemler* (1.Baskı). Pegem Akademi Yayıncılık.
- Gençtürk, A., & Memiş, A. (2010). İlköğretim okulu öğretmenlerinin öz-yeterlik algıları ve iş doyumlarının demografik faktörler açısından incelenmesi. *İlköğretim Online*, 9(3), 1037-1054.
- Gül, Ş., & Sözbilir, M. (2015). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi. *Eğitim ve Bilim*, 40(178), 85-102. <https://doi.org/10.15390/EB.2015.4070>

- Güler, G., & Ayan, C. (2020). Review of attitude scales developed in Turkey between 2002-2018 regarding the scale development process. *Journal of Faculty of Educational Sciences*, 53(3), 839-863. <https://doi.org/10.30964/auebfd.658488>
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104-121. <https://doi.org/10.1177/109442819800100106>
- Knafl, G. J., Dixon, J.K., O'Malley, J.P., Grey, M., Deatrick, J.A., Gallo, A., & Knafl, K.A. (2012). Scale development based on likelihood cross-validation. *Statistical Methods in Medical Research*, 21(6), 599-619.
- 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.
- Özdamar, K. (2003). *Modern bilimsel araştırma yöntemleri*. Kaan Kitapevi.
- Özoğlu, S. Ç. (2007). *Eğitimde rehberlik ve psikolojik danışma*. (3.Baskı). Ankara Üniversitesi Eğitim Bilimleri Yayınları.
- Simms, L.J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, 2(1), 414-433.
- Slavec, A., & Drnovsek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review*, 14(1), 39-62.
- Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Nobel Yayınevi.
- Topu, F. B., Baydaş, Ö., Turan, Z., & Gökteş, Y. (2013). Öğretim teknolojisi araştırmalarında geçerlik ve güvenirlik önlemleri. *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*. XI. Ulusal Fen bilimleri ve Matematik Eğitimi Kongresi'nde sunulan sözlü bildiri, Adana.
- Uçan, A. (2005). *Müzik eğitimi temel kavramlar-ilkeler-yaklaşımlar ve Türkiye'deki durum* (3.Baskı). Evrensel Müzikevi.
- 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.
- Yılmaz, A. (2018). *Fen bilgisi öğretmen yetiştirme programlarında kalite standartlarının belirlenmesi: Ölçek geliştirme ve uygulama çalışması*. [Doktora Tezi]. Kastamonu Üniversitesi Fen Bilimleri Enstitüsü, Kastamonu.
- Yılmaz, A. (2021). The effect of technology integration in education on prospective teachers' critical and creative thinking, multidimensional 21st century skills and academic achievements. *Participatory Educational Research*, 8(2), 163-199. <http://dx.doi.org/10.17275/per.21.35.8.2>
- Yurdugül, H. (2005). *Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması*. 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. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, Özel Sayı 2*, 264-271.