Research Article

Psychometric Characteristics of Written Response Instruments Used in Postgraduate Theses Completed in Special Education

Gamze Sarikas [iD][1], **Safiye Bilican Demir** [iD][2,*]

[1]Serkan Argın Secondary School, Bursa, Turkey
[2]Faculty of Education, Kocaeli University, Kocaeli, Turkey

Abstract: The purpose of this study is to examine the written response instruments used in postgraduate theses completed in the field of special education in Turkey between 2015 and 2018 and explore the psychometric profiles of these instruments. In the study, a total of 137 master's theses and 37 dissertations were reviewed using the Data Collection Instrument Review Form. Categorical and frequency analyses were used in the analysis of the data. Also, the relevant categories were discussed by citing remarkable examples of errors or deficiencies. According to the research findings, a total of 387 written response instruments were used in the theses reviewed. Most of the written response instruments were developed by the researchers themselves and of these instruments, the most frequently used were the personal information forms and scales. In the theses, most of the written response instruments were not introduced or only partly introduced and the validity and reliability of these instruments were either not reported or only partly reported. The results of the research showed that there were crucial deficiencies and errors in reporting the basic methodological information about the written response instruments used in the theses and this situation was repeated in the theses. In parallel with the results of the research, the related problems and their causes were discussed, and suggestions offered.

1. INTRODUCTION

The aim of scientific research is to produce knowledge. In this process, answers to the researched problem are sought in accordance with standard scientific principles. The production of scientific knowledge takes place only when the stages of scientific research methods are carried out completely. Universities are one of the institutions that produce and share scientific information. Universities support social development by producing information and technology through postgraduate studies and by providing qualified human resources. Theses produced as a result of postgraduate studies are of great importance in the development of a field. The theses put forward specific solution suggestions for a research problem in accordance with the scientific process steps. As such, theses show that the prospective researcher possesses the knowledge and the proficiency to carry out independent research and can produce scientific information that will contribute to the field of study (Tavşancıl et al., 2010). In addition to providing scientific standards, these research reports are valuable in that they are reviewed by

a scientific jury. Therefore, postgraduate theses play a key role in the growth and development of a discipline (Evrekli et al., 2011). In this regard, it is clear that development and change in any field is closely related to scientific research in that field (Seçer et al., 2014).

The increase in the number of universities in Turkey in recent years has particularly resulted in an increase in postgraduate education, which in turn has led to a quantitative increase in scientific research in education (Özenç & Özenç, 2013). The same is true of special education (Diken et al., 2009). Both the increase in the number of students receiving special education and the legal regulations relating to the field are seen as factors in the increase in scientific studies in the field of special education. According to the World Health Organization (WHO), approximately 12% of individuals in the 6-18 age group have special needs. Some sources put this figure as high as 14% (Metin, 2012). The number of students receiving special education services has increased by 585% from 2002 to 2013 (Melekoğlu, 2014). The increase in the number of students with special education needs has led to an increase in graduate programs training teachers to work in this field resulting in an increase in the number of preservice teachers graduating from these programs and more teaching staff working in universities' special education departments (Ağca, 2014). The importance of postgraduate programs not only in terms of producing scientific information but also in terms of training teachers who will run graduate and postgraduate programs in special education is starting to emerge. As a result, the developments in various aspects of special education are having a positive effect on the increase in scientific studies made in this field.

The quantitative increase in postgraduate studies both in the field of special education and in other fields of education can be considered a positive step for the development of the related field. However, the extent to which knowledge obtained from studies contributes to science and how scientific this knowledge is will always be a topic of debate. Knowledge can be produced through scientific studies in any field, but not all information may be truly scientific (Benligiray, 2009). In this respect, the quantitative increase in scientific studies is not always the guarantee of qualitative development. At this point, the method used to obtain the results obtained from scientific research becomes at least as significant as the results. Therefore, to produce scientific knowledge, the steps of the scientific process must be carried out completely. In this way, scientifically sound information not only reveals the facts, but it also allows scientific debate to continue by being a point of reference for new studies. However, it is known that researchers working in social sciences in Turkey experience important problems particularly when it comes to methodology (Köklü & Büyüköztürk, 1999).

Some of the methodology-related problems can be addressed in the context of data collection instruments. Researchers use various data collection techniques to obtain information about the subject of interest. Which data collection technique the researcher will use varies depending on the research problem, the nature of the data, or the source of the data. Although there are different classifications in the literature; in general, data collection approaches can be classified as written response instruments, interview, observation, available data, and document analysis (Büyüköztürk et al., 2016; Karasar, 2016; Tavşancıl et al., 2010). This study discusses the written response instruments that are used frequently in research. In the written data collection approach, communication between the researcher and the participant is made in writing and the researcher may use various data collection instruments such as questionnaire, scale, test, or inventory to collect data. No matter what type of data collection instrument the researcher uses, it is expected that this instrument's characteristics such as purpose, item number, and scoring format, etc. be reported in a scientifically appropriate manner. At the same time, the psychometric properties of the results/scores obtained from this data collection instrument should be at the desired level. Otherwise, the scientific validity of the data colection instrument and the results obtained using it will be regarded with suspicion. For the accuracy of the data

obtained using the collection instruments to be satisfactory, two fundamental characteristics are required and these are known as "reliability," which is an evidence of the stability of the results of the scores obtained from the data collection instruments, and "validity," which is an evidence of the degree to which the instrument is able to measure the characteristics it is supposed to measure (Horst, 1966). However, related studies show that although articles or theses in the field of social sciences have been published, they are not error-free and that they even contain deficiencies and errors, especially in terms of research methods (Başol & Akın, 2006; Tavşancıl et al., 2010).

Today, many different scientific studies are carried out in the field of education. When considering the theoretical or practical effects of these studies, it is necessary to classify them, determine the emerging trends, and evaluate their results (Kutluca & Demirkol, 2016; Varışoğlu et al., 2013). Staton and Wulff (1984) state that the most suitable way to do this is to review the studies in any given field periodically. These types of review can act as road maps for researchers still unfamiliar with the terminology associated with the scientific method in terms of methods used and topic selection (Cohen et al., 2007). In addition, the results of this kind of research may be considered valuable in terms of guiding research in the related field, in saving researchers' time, and facilitating access to research information. This is because an excess of work done in a field can sometimes create problems. For example, researchers who want to do research in a field often find it difficult to access all the studies done in that field or they spend more time accessing them (Göktaş et al., 2012). In this context, revealing the content and meta-analysis results of studies made in a particular field or topic by reviewing them at regular intervals makes it easier for researchers to assess the latest situations regarding their fields (Karadağ, 2009; Lee et al., 2009).

It is noteworthy that many studies have been conducted in recent years with the aim of identifying developments in different disciplines so as to determine research trends. For example; scientific studies in such fields as educational sciences (Arık & Türkmen, 2008; Doğan & Tok, 2018; Erdem, 2011; Şenyurt & Özer-Özkan, 2017; Tavşancıl et al., 2010; Yalçın, 2016), educational technologies (Alper & Gülbahar, 2009; Tosuntaş et al., 2019), curriculum (Ozan & Köse, 2014), education management (Aydın & Uysal, 2011; Turan et al., 2014), preschool education (Yılmaz & Altınkurt, 2012), science and mathematics education (Çiltaş et al., 2012; Kutluca et al., 2018; Yaşar & Papatğa, 2015) have been examined and general trends have been revealed across a very broad spectrum including research subject, research model, target group, data collection instruments, data analysis techniques, publication year, and number of authors.

As in the fields mentioned above, it is seen that trend studies have also been made in the field of special education (Aslan & Özkubat, 2019; Çoşkun et al., 2014; Demirok et al., 2015; Diken et al., 2008; Diken et al., 2016; Doğru et al., 2015; Güner-Yıldız et al., 2016; Küçüközyiğit et al., 2016; Özkubat et al., 2014; Ünlü et al., 2020; Tiryaki, 2017; Tiryakioğlu, 2014). For example, with the aim of identifying trends in special education, Ünlü et al. (2020) investigated doctoral dissertations in special education in terms of various variables. The findings obtained at the end of the research have exposed that the subject of intellectual disability was studied most in the thesis, "single subject design" was preferred mostly and a large part of the thesis was completed in Anadolu University. Also, according to the findings it has been revealed that the number of thesis in special education has been increasing in late years and more than half of all thesis were completed after 2011. Aslan and Özkubat (2019) reviewed papers published in the booklets of national congresses on special education held in Turkey between 2007 and 2017. Within the scope of the study, 1,742 papers were given a content analysis review looking at year, number of authors, sample group, research model, data collection instruments, data analysis method, and research topic categories. The results of the review showed that the

majority of the papers have one or two authors, the research topics were concentrated in inclusive education, a large number of studies used the descriptive screening model as their method, data were mostly collected by interview, and descriptive analysis was used for data analysis.

In another analysis, Tiryaki (2017) reviewed theses completed in the field of special education between the years 2000 and 2006 in terms of year of completion, institution, research subject, research model, target group, and data collection instruments. The prominent results of that review showed that in the years in question a large proportion of the theses focused on special education and language skills, there was a high proportion of qualitative research, and that data were collected using observation and interview techniques. Küçüközyiğit et al. (2016) conducted a content analysis review of 155 theses completed in the field of education for the visually impaired. That study's conclusions emphasized that there was limited study in the related field at doctoral degree and that most of the theses were descriptive and used the single-subject research model. Yıldız et al. (2016) reviewed 113 articles about special education published in journals. The results of that study showed that most of the reviewed articles had one author and that the single-subject research method was frequently preferred. Another study (Demirok et al., 2015) reviewed 400 articles published in international journals in the context of various variables such as subject, research type and sample group. The findings of that study showed that most papers were published in 2012, most of the publications had two authors, most of the articles were made using students with physical disabilities, and that experimental methods were used in more than half the articles. Studies by Çoşkun et al. (2014) and Özkubat et al. (2014) aimed to identify the research trends relating to completed theses in the field of special education. To this end, theses were assessed in terms of research field, topic, method, and pattern used, sample selection and size, data collection, and analysis. When the study's findings are analyzed, it can be seen that approximately 80% of the theses are master's theses, the topics were mostly related to skill teaching, single-subject experimental patterns are the most commonly used research pattern, most studies were quantitative, observation was the most commonly used data collection instruments, and that descriptive analysis techniques were frequently preferred.

In the studies summarized above, the scientific studies in the field of special education were subjected to a content analysis review of topic, quality and quantity, and the methods and techniques used, and an attempt was made to determine the direction of trend. For the field of special education, which is closely related to the other disciplines in the education, these kinds of studies are valuable for the development of this field. However, there is also a need for research that looks at the problem of method in studies in the field of special education and that makes a detailed psychometric review of the problem. Although there are studies that reveal the general trends in data collection instruments used in the field of special education, no studies have been found that make a detailed review to determine if these instruments meet basic psychometric standards. The role of postgraduate theses produced in any field in the development of that field is clear. At this point, the results obtained from this study are valuable in that they provide information on the psychometric properties of the data collection instruments used in postgraduate studies published in the field of special education.

This study looks at the methodology problem seen in the field of special education in terms of data collection instruments and it examines the written response instruments used in postgraduate theses completed in the field of special education as well as the psychometric characteristics of these instruments.

2. METHOD

2.1. Research Design

In this study, a qualitative research based on descriptive content analysis was used to describe the current situation of related theses.

2.2. Population

The population of the study consists of a total of 189 postgraduate theses comprising 152 master's theses and 37 dissertations completed in the field of special education between 2015 and 2018 and scanned at the Higher Education Council's National Thesis Center in Turkey. Since there are not many postgraduate theses in the population considering the researcher's conditions, sampling was not done and, instead, all the theses in the population were included in the scope of the study. The distribution of the theses in the population by distribution by years is given in Table 1.

Table 1. *Distribution by year of theses*

Year	Master's Thesis		Dissertation
	f	%	f
2015	41	26.97	16
2016	48	31.58	8
2017	35	23.03	8
2018	28	18.42	5
Total	152	100.00	37

According to Table 1, more postgraduate theses were completed in 2015 and 2016 than in the other years. A total of 41 (26.97%) of master's theses and 16 (43.24%) of dissertations were completed in 2015; 48 (31.58%) of the master's theses and eight of the dissertations were completed in 2016.

2.3. Data Collection Instruments

The study's data were obtained using the "Data Collection Instrument Review Form" for the written response instruments of the thesis review form developed by Tavşancıl et al. (2010). In this form, written response instruments were reviewed in detail under the categories of development-adaptation status, category (questionnaire, scale, and achievement/ability test), data collection instrument presentation status, development-adaptation steps, and evidence for the instruments' validity and reliability. In the thesis review form, the development steps for the researcher-developed data collection instruments were checked to see if they had been followed completely and the following steps for the questionnaire, the achievement/ability test, and the scale were considered (Tavşancıl et al., 2010):

For the questionnaire developed by the researcher:

1) Literature review
2) Review of the same or similar data collection instruments
3) Item genaration
4) Seeking expert opinion and explaining the experts' characteristics
5) Pilot study
6) Performing item analyses and/or qualitative analyses (clarity of items, etc.)
7) Deciding the final version of the questionnarie

For the achievement/ability test developed by the researcher:

1) Literature review

2) Review of the same or similar data collection instruments
3) Preparing a test blueprint
4) Item genaration
5) Seeking expert opinion and explaining the experts' characteristics
6) Pilot study
7) Performing item analyses and/or qualitative analyses (clarity of items, etc.)
8) Deciding the final version of the achievement/ability test

For the scale developed by the researcher:

1) Literature review
2) Review of the same or similar data collection instruments
3) Composition work (Creating items by examining a student's attitude, feelings, thoughts, etc. by getting the student to write a composition.)
4) Content analysis
5) Item genaration
6) Seeking expert opinion and explaining the experts' characteristics
7) Pilot study
8) Performing item analyses and/or qualitative analyses (clarity of items, etc.)
9) Deciding the final version of the scale

The presentation status of the questionnaire, achievement/ability test, and scale developed by another researcher was reviewed using the following steps (Tavşancıl et al., 2010):

1) Who developed the data collection instrument?
2) When was the data collection instrument developed?
3) For which target group was the data collection instrument developed
4) The purpose for which the data collection instrument was developed
5) Number of questions in the data collection instrument
6) The structure of the data collection instrument (graded, categorically scored, etc.)
7) How the data collection instrument is rated

The following steps were considered in the presentation of data collection instruments adapted by other researchers (Tavşancıl et al., 2010).

1) Adapted by whom
2) Adapted when
3) Adapted for which target group
4) Adapted for what purpose
5) Number of questions in the data collection instrument
6) Structure of the data collection instrument
7) Scoring method

In addition, in the thesis review form, the validity and reliability coefficients obtained for data collection instruments were determined and it was checked to see whether or not these coefficients were appropriate to the structure of the data collection instrument and whether the level was high or low. When coding for this, coefficients of 0.70 and above are considered sufficient (Tavşancıl et al., 2010).

To determine the reliability of the review form, the agreement between the coding made by different coders was 85%; the agreement between codes made by the same encoder at different times was calculated as 95% (Tavşancıl et al., 2010). Within the scope of this research, the agreement between the coding made by two different coders for the review form was calculated as 97% and the agreement between the coding made by the researcher at two different times was 86%.

2.4. Data Analysis

Categorical and frequency analyses were used in the analysis of the data. In categorical analysis, the message is first divided into units, and then these units are grouped into categories according to specific criteria. In frequency analysis, the frequency of occurrence of units and elements is determined numerically (Bilgin, 2006). Also, theses in the relevant categories are discussed by citing remarkable examples of errors or deficiencies.

3. RESULTS

The findings are given according to the categories in the review form. Firstly, the use of the written data collection technique in postgraduate theses was examined and the findings are given in Table 2.

Table 2. *Distribution for use of written data collection technique*

Data Collection Techniques	Master's Thesis	f	%	Dissertation f
Written	Used	133	87.50	31
Response	Unused	19	12.50	6
Instrument	Total	152	100.00	37

According to Table 2, the written data collection technique was used in 133 (87.50%) of the reviewed master's theses and in 31 of the dissertations. The written response instruments were used in a total of 272 master's theses and 115 dissertations.

The written response instruments were reviewed on the following bases: developed by the researcher, developed by another researcher, adapted by the researcher, and adapted by another researcher; the distribution is given in Table 3.

Table 3. *Distribution of written data collection instruments by development and adaptation*

Measuring Instrument Development/ Adaptation Status	Master's Thesis f	%	Dissertation f	%
Researcher Developed	161	59.19	68	59.13
Researcher Adapted	-	-	-	-
Developed Instrument Used	52	19.12	17	14.78
Adapted Instrument Used	59	21.69	30	26.09
Total	272	100.00	115	100.00

When Table 3 is examined, it can be seen that 161 (59.19%) of the written data collection instrument used at the master's degree were developed by the researcher, 52 (19.12%) were developed by other researchers, and 59 (21.69%) were adapted by other researchers. There are no adapted instruments used by the researcher. At doctoral degrees, 68 (59.13%) of the instruments were developed by the researcher, 17 (14.78%) were developed by other researchers and 30 (26.09%) were adapted by other researchers. Among the instruments used at doctoral degrees, there are no adapted instruments used by the researcher. At both degrees, the written response instruments used by the researcher were frequently those developed by the researcher.

3.1. Findings for Written Response Instruments Developed by Researcher

3.1.1. *Category*

Various data collection instruments have different properties in terms of measurement technique and the steps that need to be followed when developing them differ. Taking this into consideration, the researcher-developed instruments were groups in three categories, namely, "questionnaire," "achievement/ability test," and "scale", and they were examined in accordance with the instrument development steps required by their respective categories. The data collection instruments that are not included in these categories are reviewed under the "other" category. These include checklists, forms for the social validity of the research, reinforcement determination lists, evaluation forms, personal information forms, and similar data collection instruments. The distribution by category of the written response instruments developed by the researcher is presented in Table 4.

Table 4. *Distribution by category of written response instruments*

Category	Master's Thesis		Dissertation	
	f	%	f	%
Questionnaire	36	22.36	7	10.30
Achievement-Ability Test	7	4.35	8	11.76
Scale	12	7.45	4	5.88
Other	106	65.84	49	72.06
Total	161	100.00	68	100.00

According to Table 4, of the 161 data collection instruments developed by the researcher at master's degrees, 36 (22.36%) were questionnaires, seven (4.35%) were achievement/ability tests, 12 (7.45%) were scales, and 106 (65.84%) were other. Of the 68 data collection instruments developed by the researcher at doctoral degrees, seven (10.30%) were questionnaires, eight (11.76%) were achievement/ability tests, four (5.58%) were scales, and 49 (72.06%) were other. The distribution of the development steps as reported or not for the data collection instruments developed by the researcher is presented in Table 5.

Table 5. *Distribution for reporting of development steps*

Category	Development Steps	Master's Thesis f	Dissertation f
Questionnaire	Reported	6	2
	Not reported	30	5
	Total	36	7
Achievement-Ability Test	Reported	1	5
	Not reported	6	3
	Total	7	8
Scale	Reported	12	4
	Not reported	-	-
	Total	12	4

When Table 5 is examined, it can be seen that of all the instruments developed by the researcher at master's level, development steps were reported for six of the 36 questionnaires, one of the seven achievement/ability tests, and all 12 scales; at doctoral level, development steps were reported for one of the seven questionnaires, five of the eight achievement/ability tests, and all four scales. Accordingly, it is remarkable that the development steps for most of the data collection instruments developed by the researcher at both levels were not reported. The

distribution of complete/incomplete development steps for those data collection instruments where the development steps have been stated is given in Table 6.

Table 6. *Distribution of complete/incomplete development steps*

Category	Development Steps	Master's Thesis f	Dissertation f
Questionnaire	Complete	1	-
	Incomplete	5	2
	Total	6	2
Achievement-Ability Test	Complete	-	-
	Incomplete	1	5
	Total	1	5
Scale	Complete	3	1
	Incomplete	9	3
	Total	12	4

According to Table 6, of the 12 scales with reported development steps at master's level, three are complete; the development steps for only one scale at doctoral level are complete. However, six of the master's level questionnaires, nine of the scales, and one achievement/ability test were found to have incomplete development steps. In dissertations, two of the questionnaires, five of the achievement/ability tests, and three of the scales were found to have incomplete development steps. Incomplete development steps are a situation that is frequently encountered in all categories of instruments at both degrees.

When the data collection instruments determined to be missing in the development steps are reviewed, for two surveys at master's level and five at doctoral level, the most common missing steps are "item generation", "stating the rate of feedback", "conducting item analyses and/or qualitative analyses on the data obtained from the application, determining the psychometric properties", and "stating whether or not monitoring was carried out"; while the most common missing steps in one achievement and ability test level at master's level and five achievement and ability tests at doctoral level were determined as "preparing a test blueprint", "establishing a pool of items", "seeking expert opinion and explaining the characteristics of experts", and "pilot study". It was determined that the most common missing steps in seven scales at master's level and three at doctoral level were the "composition work", "content analysis", and "establishing a pool of items" steps.

3.1.2. *Proof of validity*

The first examination of the psychometric properties of the written response instruments developed by the researcher was carried out using validity prediction methods. Whether or not proof of validity for the results obtained by researcher-developed data collection instruments in the relevant category was stated and what kind of proof of validity was presented were examined and the findings are given in Table 7. When Table 7 is examined, it can be seen that while proof of validity is not provided for most of the questionnaires or any of the achievement/ability tests developed by the researcher at master's level, proof of validity is provided for most of the scales; and at doctoral level it can be seen that proof of validity is stated for two questionnaires, four achievement/ability tests, and two scales. At both degrees, proof of validity is not reported for questionnaires. At doctoral level, proof of validity was reported for four achievement/ability tests; the validity prediction method used for these was construct validity based on exploratory factor analysis (EFA) only; in addition, in only one achievement/ability test, were EFA and confirmatory factor analysis used together. The

distribution of validity prediction methods used in researcher-developed scales is given in Table 8.

Table 7. *Distribution for reported proof of validity by instrument category*

Category	Proof of Validity	Master's Thesis f	Dissertation f
Questionnaire	Reported	1	2
	Not reported	35	0
	Total	36	2
Achievement-Ability Test	Reported	-	4
	Not reported	7	4
	Total	7	8
Scale	Reported	11	2
	Not reported	1	2
	Total	12	4

Table 8. *Distribution of validity prediction methods used in scales developed by researcher*

Validity prediction Methods		Master's Thesis f	Dissertation f
Content Validity	Reported	9	1
	Not reported	3	-
	Total	12	1
Construct validity	Reported	7	1
	Not reported	4	-
	Total	11	1
Criteria Based Validity	Reported	1	-
	Not reported	10	1
	Total	11	1

When Table 8 is examined, it can be seen that at master's level, structure and content validity are mainly used for scales; while at doctoral level, content and construct validity were used for one scale. At both levels, expert opinion was used for content validity and EFA was frequently used for construct validity.

3.1.3. *Proof of reliability*

The second examination of the psychometric properties of the written response instruments developed by the researcher was carried out using reliability prediction methods. The distribution of reported proof of reliability for these instruments developed by the researcher is given in Table 9.

Table 9. *Distribution for reported proof of reliability*

Category	Proof of Reliability	Master's Thesis f	Dissertation f
Achievement-Ability Test	Reported	-	4
	Not reported	7	4
	Total	7	8
Scale	Reported	10	1
	Not reported	2	3
	Total	12	4

According to Table 9, proof of reliability is reported for 10 of 12 scales at master's level; while at doctoral level, proof of reliability is reported for four of eight achievement/ability tests and one of four scales. It was observed that no proof of reliability was reported questionnaires at both levels and for seven achievement/ability tests at master's level.

In cases where researcher-developed written response instruments were used, instances were observed where proof for the instrument was reported using only one validity prediction method, as were instances where proof was reported using multiple validity prediction methods. Taking this into consideration, every single validity prediction method reported for the achievement/skill tests and scales was examined separately. The distribution of predicted reliability type for the achievement/ability test at doctoral level is given in Table 10.

Table 10. *Distribution of achievement/ability test reliability coefficients by type and level*

Reliability Prediction Method		Dissertaion f
Alpha Reliability	Not predicted	3
	Appropriate and high level	1
	Total	4
Test-Retest Reliability	Not predicted	3
	Appropriate and high level	1
	Total	4
KR-20 Reliability	Appropriate and high level	4
	Total	4

According to Table 10, Cronbach's Alpha reliability, test-retest, and KR-20 reliability were calculated for the achievement/ability tests developed by the researcher at the doctoral level These calculated reliability coefficients were found to be appropriate to the structure of the measuring instrument used and the predicted value was also high (0.70 and above). The KR-20 coefficient was used most frequently as proof of reliability for achievement/ability tests.

The distribution of reliability coefficients in the scales developed by the researcher by predicted status and predicted reliability type and level is given in Table 11.

Table 11. *Distribution of reliability coefficients in the scales by type and level*

Reliability Prediction Method		Master's Thesis f	Dissertation f
Alpha Reliability	Not predicted	1	3
	Appropriate and high level	10	1
	Appropriate and low level	1	-
	Total	12	4
Test-Retest Reliability	Not predicted	7	3
	Appropriate and high level	5	1
	Total	12	4
Split-half Reliability	Not predicted	10	4
	Appropriate and high level	2	-
	Total	12	4

According to Table 11, Cronbach's Alpha, test-retest, and split-half reliability were calculated for scales at master's level. While Cronbach's Alpha is the most common value calculated for scales, it was observed that the alpha value calculated for one scale was low. At doctoral level,

Cronbach's alpha and test-retest reliability were calculated for the scales. These coefficients were found to be high and appropriate to the scale structure.

3.2. Findings on Written Response Instruments Developed by Other Researchers

3.2.1. *Category*

It was determined that 52 data collection instruments were developed by other researchers at master's level and 16 at doctoral level. The first criterion taken into consideration when evaluating the presentation of these instrument in postgraduate theses is the instrument's category. The distribution of data collection instruments developed by other researchers by categories is given in Table 12.

Table 12. *Distribution of data collection instruments developed by other researchers by category*

Category	Master's Thesis		Dissertation
	f	%	f
Achievement-Ability Test	15	28.85	4
Scale	29	55.77	3
Other	8	15.38	10
Total	52	100.00	17

When Table 12 is examined, it can be seen that 15 (28.85%) of the 52 data collection instruments at master's level are achievement/ability tests, 29 (55.77%) are scales and eight (17.30%) are in the other data collection instruments category; while at doctoral level, four of the instruments are achievement/ability tests, three of them are scales and 10 of them are in the other category. It was also determined that no questionnaire developed by another researcher was at the master's or doctorate degrees.

Another examination of the data collection instruments developed by other researchers looked at the how the instrument was introduced. The instrument categories were examined to see if the information that needs to be presented when introducing the instrument was reported. The distribution for the introduction of data collection instruments developed by other researchers in the "achievement/ability test" and the "scale" categories is given in Table 13.

Table 13. Distribution of introduction status by instrument category

Category	Introduction Status	Master's Thesis	Dissertation
		f	f
Achievement-Ability Test	Introduced	3	1
	Not Fully Introduced	10	2
	Not introduced	2	1
	Total	15	4
Scale	Introduced	15	1
	Not Fully Introduced	12	1
	Not introduced	2	1
	Total	29	3

According to Table 13, at master's level, 10 achievement/ability tests were fully introduced while three were not fully introduced or not introduced at all. Of the scales, 15 were introduced, 12 were not fully introduced and two were not introduced at all. It can be seen that at doctoral level, one of the four achievement/ability tests was introduced, two were not fully introduced, and one not introduced at all; while of the three scales, one was introduced, one was not fully introduced, and one not introduced at all.

It was determined that at master's level, the most frequently observed deficiency in the intrduction of achievement/ability tests developed by other researchers was due to information about the structure of the instrument (graded scale, etc.) not being given. At doctoral level, it was found that information about the number of questions in the instrument, the structure of the instrument, and the instrument's scoring method was not given.

3.2.2. *Proof of validity*

The written response intruments developed by other researchers were also examined in terms of the methods used for predicting the validity. Information reported in the theses with respect to validity prediction methods and proof of validity was examined under two categories, namely, "Original proof of validity" and "Proof of validity as reported in the study". Whether or not original proof of validity was given for data collection instruments developed by other researchers was examined and the the findings are given in Table 14.

Table 14. *Distribution of reported original proof of validity*

Category	Original Proof of Validity	Master's Thesis f	Dissertation f
Achievement-Ability Test	Reported	10	1
	Not reported	5	3
	Total	15	4
Scale	Reported	14	2
	Not reported	15	1
	Total	29	3

According to Table 14, at master's level, original proof of validity was given for 10 of achievement/ability test. It was determined that the most frequently cited proof was construct validity, content validity, and criterion-based validity. At this level, original proof of validity was given for 14 of the scales and this evidence was frequently based on construct validity, criterion-based validity, and content validity. At doctoral level, original proof of validity was given for one achievement/ability test and two scales. This evidence was often based on construct validity.

Presentation of proof of validity for data collection instruments developed by other researchers within the scope of the study was examined and it was determined that absolutely no proof of validity was reported for the achievement/ability tests within the scope of the study; and that only at master's level was construct validity proof of validity reported within the scope of the study for two scales.

3.2.3. *Proof of reliability*

Information regarding proof of reliability reported in postgraduate theses was examined as "Original proof of reliability" and "Proof of reliability reported in the study". Presentation of original proof of reliability in theses where an instrument developed by someone else was used was examined and the distribution is given in Table 15.

Table 15. *Distribution of reported original proof of reliability*

Category	Original Proof of Reliability	Master's Thesis f	Dissertation f
Achievement-Ability Test	Reported	10	2
	Not reported	5	2
	Total	15	4
Scale	Reported	22	2
	Not reported	7	1
	Total	29	3

According to Table 15, at master's level original proof of reliability was reported for 10 achievement/ability tests and 22 scales, while at doctoral level it was reported for just two achievement/ability tests and two scales. It was found that the Cronbach's Alpha coefficient, test-retest, split-half reliability, and inter-rater reliability coefficient were the ones most frequently calculated for achievement/ability tests developed by other researchers at master's level and that their level was high. At doctoral level, it was observed that original proof of reliability based on test-retest reliability was obtained and that these values were high. The Cronbach's alpha coefficient, test-retest and two-half reliability evidence were given as proof of original reliability in all 22 scales used at the master's level, and the values obtained from them were found to be high. However, the Cronbach's Alpha reliability coefficient calculated for the two instruments was found to be low. At doctoral level, the original proof of reliability was calculated based on Cronbach's alpha, test-retest, split-half, and KR - 20 reliability.

In addition, proof of reliability for written response instruments developed by other researchers and obtained within the scope of postgraduate theses was examined. While the reliability coefficient calculated for five scales developed by other researchers at master's level was high, it was determined that the reliability coefficient calculated for two scales was low. At doctoral level, it was seen that no proof of reliability was reported in the relevant theses within the scope of the study.

3.3. Findings on Written Response Instrument Adapted by Other Researchers

3.3.1. *Category*

It was determined that 59 written response instruments at master's level and 30 instruments at doctoral level were adapted to Turkish culture by other researchers. The distribution of instruments adapted by other researchers by the categories of "questionnaire," "achievement/ability test," and "scale" is given in Table 16.

Table 16. *Distribution of adapted instruments by category*

Category	Master's Thesis f	Dissertation f
Achievement-Ability Test	8	2
Scale	36	24
Other	13	4
Total	59	30

In Table 16, it can be seen that eight of the written response instruments at master's level were achievement/ability tests, 36 were scales, and 13 fell into the "other" category; while at doctoral level, two were achievement/ability tests, 24 were scales, and four were in the "other" category.

Another examination, this time of written response instruments adapted by other researchers, sought whether or not these instruments had been introduced and if so whether or not the reported information was complete. The introduction status of the data collection instruments adapted by other researchers was examined and the the findings are given in Table 17. When Table 17 is examined, it can be seen that three achievement/ability tests and 27 scales adapted by other researchers were introduced but that four achievement/ability tests and nine scales were not fully introduced at master's level. It was found that one achievement/ability test and 20 scales adapted by other researchers were introduced but that one achievement/ability test and three scales had incomplete introduction information at doctoral level.

Table 17. *Distribution of introduction status of adapted instruments*

Category	Introduction Status	Master's Degree f	Dissertation f
Achievement/Ability Test	Introduced	3	1
	Not Fully Introduced	4	1
	Not introduced	1	-
	Total	8	2
Scale	Introduced	27	20
	Not Fully Introduced	9	3
	Not introduced	-	1
	Total	36	24

At master's level, it was determined that incomplete information was given for the achievement/ability test, most frequently in the "number of questions in the instrument", "structure (graded scale etc.)", and the "scoring method" areas. At doctoral level, incomplete information was given only for "instrument structure (graded scale etc.)". For scales, it was found that at master's level, incomplete information was given for "adapted by whom", "adapted when", and "scoring method"; and that at doctoral level incomplete information was given for "instrument structure (graded scale etc.)", "adapted by whom", "adapted when", "number of questions in the instrument", and "scoring method".

3.3.2. *Proof of validity*

The first examination of the psychometric properties of written response instruments adapted by other researchers was made using validity prediction methods. For these instruments, the original proof of validity reported by the researcher who developed the instrument, the proof of validity obtained in his/her own studies by the researcher who adapted the instrument to Turkish culture, and the proof of validity reported in their studies by researchers who used the adapted instrument all needed to be reported. Accordingly, the information obtained was presented as "Original proof of validity," "Proof of validity reported by researchers who adapted the instrment to Turkish culture," and "Proof of validity reported in the study". Whether or not original proof of validity for written response instruments adapted by other researchers was reported was examined and the resulting distribution is given in Table 18.

Table 18. *Distribution of original validity prediction methods in adapted scales*

Original Validity Prediction Method		Master's Thesis f	Dissertation f
Construct validity	Reported	7	2
	Not reported	7	-
	Total	14	2
Predictive Validity	Reported	3	-
	Not reported	11	2
	Total	14	2
Criteria Based Validity	Reported	3	-
	Not reported	11	2
	Total	14	2
Content Validity	Reported	13	-
	Not reported	1	2
	Total	14	2

According to Table 18, at master's level, construct validity was given as original proof of validity for seven scales, criterion-based and predictive validity for three scales, and content

validity for one scale. At doctoral level, construct validity was reported for the original validity of the two scales.

Whether or not proof of validity determined by researchers who took written response instruments adapted by other researchers and adapted them to Turkish culture was stated was examined and the resulting distribution is given in Table 19.

Table 19. *Distribution for reported proof of validity for adapted instruments*

Original Validity Prediction Method	Proof of Validity	Master's Thesis f	Dissertation f
Achievement/Ability Test	Reported	1	-
	Not reported	7	2
	Total	8	2
Scale	Reported	17	10
	Not reported	19	14
	Total	36	24

In Table 19, it can be seen that proof of validity determined by researchers who adapted the data collection instrument to Turkish culture was stated for one achievement/ability test and 17 scales at master's level and for 10 scales at doctoral level. It was determined that content and construct validity were used as the validity determination method for the achievement/ability test at master's level. It was found that construct validity and proof based on criterion-based validity were used for scales. In addition to this, it was seen that at doctoral level, construct validity based on factor analysis was stated for almost all the scales, while proof was reported based on criterion-based validity by using similar scales in another instrument.

It was determined that proof of validity obtained within the scope of the study was not reported for written response instruments adapted by other researchers.

3.3.3. *Proof of reliability*

The second examination of the psychometric properties of written response instruments developed by adapted by other researchers was made using reliability prediction methods. The findings are presented as "Original proof of reliability," "Proof of reliability reported by researchers who adapted the instrument to Turkish culture," and "Proof of reliability reported in the study". For data collection instruments adapted by other researchers, whether or not original proof of validity determined by researchers who developed the instrument for the original culture was stated was examined and the findings are given in Table 20.

Table 20. *Distribution of reported original proof of reliability for adapted instruments*

Category	Original Proof of Reliability	Master's Thesis f	Dissertation f
Achievement-Ability Test	Reported	1	-
	Not reported	7	2
	Total	8	2
Scale	Reported	17	6
	Not reported	19	18
	Total	36	24

According to Table 20, at master's level, original proof of reliability was reported for one achievement/ability test and 17 scales adapted by other researchers. It was determined that at doctoral level, original proof of reliability was not reported for achievement/ability tests but

was reported for just six scales. At master's level, the original reliability prediction method for one achievement/ability test was calculated using the KR - 21 coefficient.

Table 21 shows the distribution by type and level of predicted reliability for stated original reliability coefficients stated as have been predicted by the researchers who developed the instrument for scales adapted by other researchers.

Table 21. *Distribution of prediction of original reliability coefficients in adapted scaled by type and level of predicted reliability*

Original Reliability Prediction Method		Master's Thesis f	Dissertation f
Alpha Reliability	Not predicted	3	1
	Appropriate and high level	11	3
	Appropriate and low level	3	1
	No information about level	-	1
	Total	17	6
Test-Retest Reliability	Not predicted	11	6
	Appropriate and high level	6	-
	Total	17	6
KR-20 Reliability	Not predicted	17	5
	Appropriate and high level	-	1
	Total	17	6
KR - 21 Reliability	Not predicted	16	6
	Appropriate and high level	1	-
	Total	17	6

According to Table 21, Cronbach's Alpha reliability was estimated for 14 of the 17 scales as the original proof of reliability in theses at master's level and the estimated value was high for 11 scales but low for three scales. For the three scales, the relevant coefficients were not reported. Remarkably, high test-retest reliability coefficient was calculated for six of the 17 scales, while KR - 21 coefficient was calculated for one scale. It can be seen that at doctoral level, Cronbach's Alpha reliability was predicted for five of the scales of which three were found to have appropriate scale structure and high values; while one was found to have an appropriate scale structure but a low value. For one scale the relevant coefficient was stated as having been predicted but no information for this value is given and that the KR - 20 coefficient was calculated for one other scale.

In the examination of the psychometric properties of written response instruments adapted by other researchers, the reporting in postgraduate theses of the proof of reliability obtained by researchers who adapted the instrument to Turkish culture was examined. The distribution of reported proof of reliability determined by the researchers adapting the instrument to the Turkish culture is given in Table 22. When Table 22 is examined, stated proof of reliability determined by researchers who adapted the instruments to Turkish culture can be seen for four achievement/ability tests and 24 scales at master's level and for two achievement/ability tests and 13 scales at doctoral level.

Table 22. *Distribution of reported proof of validity determined by researchers who took adapted instruments and adapted them to turkish culture*

Category	Proof of Reliability	Master's Thesis f	Dissertation f
Achievement-Ability Test	Reported	4	2
	Not reported	4	-
	Total	8	2
Scale	Reported	24	13
	Not reported	12	11
	Total	36	24

Table 23 and Table 24 show the distribution of the type and level of the coefficients for achievement/ability tests and scales, respectively. According to Table 23, the reliability coefficients predicted by the researchers who adapted the instrument to Turkish culture were Cronbach's Alpha, the split-half test, and the KR - 21 at master's level and Cronbach's Alpha, test-retest, and parallel test form reliability at doctoral level. It was determined that these calculated coefficients were appropriate to the structure of the instrument and their level was high. It was determined that the reliability of the split-half test for the achievement/ability test was low. In Table 24, it is seen that Cronbach's Alpha test re-test, split-half, and KR - 20, and parallel test reliability were used respectively for scales at master's level. For scales at doctoral level, Cronbach's alpha, test-retest, split-half, and parallel test reliability were used. Of these scales, it was found that the Cronbach's Alpha reliability coefficient was low for two scales at master's level and one at doctoral level and that at doctoral level, the test-retest and split-half reliability coefficients for one scale were low. It was noted that the KR - 20 coefficient, which is not appropriate to the structure of the scale, was calculated as proof of reliability at master's level. The majority of the scales used at both levels were found appropriate to the structure of the instrument and to have high proof of reliability.

Table 23. *Distribution of achievement/ability test adapted to turkish culture by reliability coefficient prediction and type and level of predicted reliability*

Original Reliability Prediction Method		Master's Thesis f	Dissertation f
Cronbach's Alpha Reliability	Not predicted	7	1
	Appropriate and high level	1	1
	Total	8	2
Test-Retest Reliability	Not predicted	8	1
	Appropriate and low level	-	1
	Total	8	2
Split-half Reliability	Not predicted	7	1
	Appropriate and low level	1	1
	Total	8	2
KR - 21 Reliability	Not predicted	7	2
	Appropriate and high level	1	-
	Total	8	2
Parallel Test Reliability	Not predicted	8	1
	Appropriate and high level	-	1
	Total	8	2

Table 24. *Distribution by reliability coefficients type and level of predicted reliability for adapted sacels*

Reliability Prediction Method		Master's Thesis f	Dissertation f
Cronbach's Alpha Reliability	Not predicted	12	12
	Appropriate and high level	22	11
	Appropriate and low level	2	1
	Total	36	24
Test-Retest Reliability	Not predicted	25	21
	Appropriate and high level	11	2
	Appropriate and low level	-	1
	Total	36	24
KR-20 Reliability	Not predicted	34	24
	Appropriate and high level	2	-
	Total	36	24
Split-half Reliability	Not predicted	26	22
	Appropriate and high level	10	1
	Appropriate and low level	-	1
	Total	36	24
Parallel Test Reliability	Not predicted	34	23
	Appropriate and high level	2	1
	Total	36	24

When examined within the scope of the study, it was found that for instruments adapted by other researchers, at master's level, proof of reliability was reported for eight out of 36 scales and that for all of them a high Cronbach's Alpha reliability coefficient was obtained. At doctoral level, it was found that there was only one scale with reported proof of reliability and that a high Cronbach's Alpha value was obtained for this scale.

4. DISCUSSION and CONCLUSION

In this study, written response instruments used in postgraduate theses completed in the field of special education between 2015 and 2018 and their psychometric properties were examined. It was seen that most of the written data collection technique at both degrees were researcher-developed and that no researcher-adapted written response instruments were used.

The most frequently used researcher-developed written response instruments were the questionnaire at master's level and the achievement/ability test at doctoral level. Consistent with the findings of this study, it was determined that surveys and achievement tests were frequently used in the studies published in the field of educational sciences both in Turkey and abroad (Doğru et al., 2012; Erdem, 2011; Tavşancıl et al., 2010; Yalçın, 2016; Yalçın et al., 2015).

It was noted that development steps were not reported in more than half of the researcher-developed questionnaire, achievement/ability test, and scale. In a similar study, Tavşancıl et al. (2010) stated that there were significant deficiencies in reporting the development steps of the data collection instruments developed by the researchers in the relevant theses at master and doctoral level; given that this information given in a limited number of theses, it was concluded that the measurement procedures were not done with sufficient quality. In studies by Başol and Akın (2006) and Arık and Türkmen (2009), it was emphasized that in the articles they reviewed, there was not enough information about the data collection instruments used and that this situation could negatively impact the intelligibility of the studies. This indicates that the deficiencies in the introduction of data collection instruments used in studies are still seen today.

When the results of reported proof of validity in researcher-developed instruments are examined, it can be seen that proof of validity was not reported for any of the

achievement/ability tests used at master's level, that it was presented in a large majority of the scales, and that the most frequently used proof of validity for scales were construct and content validity. At doctoral level, it was observed that proof of validity for researcher-developed achievement/ability tests and scales was rarely reported. In findings relating to proof of reliability for researcher-developed instruments, it was seen that at master's level, proof of reliability was not reported for any questionnaire or achievement/ability test but that proof of reliability was reported for a large majority of the scales. Cronbach's Alpha coefficient was frequently used for predicting reliability at the master's level; the calculated coefficients were found to be high and consistent with the scale's scoring structure. Similarly, in the study conducted by Mor-Dirlik and Kula-Kartal (2016), it was stressed that Cronbach's Alpha coefficient was the most commonly used evidence of reliability in both education and psychology. It was found that proof of reliability was given for half the achievement/ability tests and only one scale used at doctoral level. For achievement/ability tests, it was determined that the KR - 20 coefficient was calculated most frequently, and these coefficients were found to be high and consistent with the instrument's scoring method. Tavşancıl et al. (2010) showed that the KR-20 reliability coefficient was frequently reported for achievement tests in postgraduate theses in the field of educational sciences.

The results of the study show that there were significant deficiencies in the validity and reliability of the written response instruments developed by the researcher. This will bring controversy about the accuracy of the results obtained from research in which data collection instruments of dubious validity and reliability were used. As it is known, validity and reliability are the basic psychometric properties required of a data collection isntruments. The meaning of the scores obtained from a data collection instrument and the lack of evidence that the instrument makes accurate measurements without confusing research variables with other variables could cause the relevant research results to become questionable for both readers and other researchers in terms of the scientific method. It was noted that although there are many methods for predicting validity for those instruments that had proof of validity presented, only proof of construct-related evidence of validity was reported. However, if validity studies are considered to be the process of collecting evidence for the accuracy of the scores obtained from the data collection instruments, it stands to reason that evidence obtained from different validation methods will contribute to the accuracy of the research results. The same is true for reliability. In related studies (Başol & Akın, 2006; Büyüköztürk & Kutlu, 2006; Tavşancıl et al., 2010), it was emphasized that the failure to present validity and reliability information for data collection instruments was the most serious methodological problem.

Another conclusion of the study was that of the data collection instruments developed by other researchers, the scale was the one used most frequently at master's level and the achievement/ability test at doctoral level. When the presentation status of the developed instruments (structure, scoring method, etc.) was examined, it was seen that at master's level in most of the achievement/ability tests and the scales, the most common missing information were *structure* and *scoring* method. At doctoral level, this includes *the number of questions, the structure of the instrument,* and *the scoring* method.

For the data collection instruments developed by other researchers, it was observed that original proof of validity for the achievement/ability test was mostly reported at master's level and that this proof of validity was construct validity based on factor analysis. At doctoral level, it was found that the achievement/ability test and the scale were those instruments for which original proof of validity was not reported. However, it was found that construct validity was reported for most of the written response instruments for which original proof of validity was reported. At master's level, it was reported that Cronbach's Alpha reliability coefficient, which has a high reliability coefficient and is appropriate to the scale structure, was the coefficient calculated the

most for achievement/ability tests and scales developed by other researchers. However, it was seen that the original reliability coefficients were presented based on test-retest, split-half, and inter-rater reliability. At doctoral level, it was noted that the frequency of reporting original proof of reliability for achievement/ability tests and scales was low. In addition, when calculating the reliability coefficient for some scales, it was noted that the KR-20 coefficient, which is not appropriate for the graded structure of these instruments, was calculated. It was determined that proof of reliability obtained within the scope of the study for achievement/ability tests and scales developed by other researchers was not reported.

In the reviewed theses, it was determined that the data collection instruments adapted by other researchers were mostly in the scale and achievement/ability test category and that most of these instruments were introduced. The original proof of validity for scales adapted by other researchers was most often found to be based on construct validity. At neither level original proof of validity was reported for adapted achievement/ability tests. Similarly, within the scope of the study, it was noted that no validation work was carried out for the adapted written response instruments. Original proof of reliability was given for scales adapted by other researchers. Cronbach's Alpha reliability was frequently used for scales at both levels. However, it was found that the KR - 21 coefficient, which is not appropriate for graded scales, was calculated at master's level and that the KR - 20 coefficient was calculated at doctoral level. When proof of reliability status obtained by researchers who adapted instruments to Turkish culture was examined, it was seen that proof of reliability was reported for most scales at both levels. It was determined that Cronbach's Alpha reliability coefficients, which are appropriate to the instrument's structure and have a high level, were used for the data collection instruments at both levels. However, it was observed that at master's level, a low Cronbach's Alpha coefficient was obtained for some scales. In addition, again at master's level, it was found that the KR - 20 reliability coefficient, which is not appropriate for determining the reliability of scales and is applied only for dichotomously scored items, was used. Similarly, in a study by Tavşancıl et al. (2010), it was seen that reliability prediction methods such as KR-20 and KR-21 coefficients, which are not appropriate for data collection instruments consisting of graded items and which can only be used when the item structure is dichotomous, were reported when predicting reliability for scales. For instruments adapted by other researchers, within the scope of the study, it was found that Cronbach's Alpha proof of reliability was presented for scales only.

In the reviewed theses, the existence of several serious repeated mistakes was noted. The most common of these repeated errors is the discrepancy between the written response instrument and its name. Some of the written response instrument used in master's theses were called *questionnaire* but it was seen that these instruments were actually scales that give total scores. For example, in the instruments called *"Frequency of Use of Phonological Awareness in Teaching Activities Questionnaire"* and *"Strengths and Difficulties Questionnaire"* it was seen that the items were scored using a five-point Likert scale able to obtain total scores. Similarly, in the study conducted by Tavşancıl et al. (2010), problems were seen that stemmed from the concepts of questionnaire and scale being used interchangeably. Another remarkable situation relating to written data collection instruments is the fact that the names in some theses are quite general and not understandable. For example, as seen in the *"Collecting Effectiveness Data"*, *"Discretionary Reinforcement Processing Criteria-Dependent Measurement Tool"* and *"Start Level Data Form"* and, likewise, the *"Productivity Data Collection Form"* and *"Start Level Sessions Form"*, it was seen that some of the written response instrument names are very general with no information given as to what structure it measures.

Another common error in the reviewed theses is related to obtaining and interpreting proof of validity. For example; for one scale, *"the findings obtained from a study of the Turkish version's*

psychometric characteristics concluded that the scale was valid and reliable", and in another study, *"work on the original version of the scale and the Turkish form presented proof of validity and reliability"* saying that the scale was valid and reliable. This shows that the researcher has incomplete or inaccurate information about how to reflect the basic psychometric properties of the data collection instruments. Some of the researchers, on the other hand, concluded that the written response instrument is valid based on the assumptions of factor analysis. For example, as proof of validity for one scale that was used, *"the KMO Barlett coefficient was applied for construct validity and was found to be 0.79",* and it was seen that the researcher accepted validity assumptions as proof of validity.

It has been determined that there are serious deficiencies in introducing the data collection instruments used in the postgraduate theses and reporting their psychometric properties. At this point, it might be a good suggestion for researchers to work on developing their research methodology and academic reporting skills and for official units to be formed where they could receive advice. In addition, a "Thesis Writing Guide" based on standards to be formulated jointly by all universities could be prepared. It is noteworthy that similar errors are repeated in the reviewed theses. In this respect, graded scoring keys or Thesis Review Forms could be developed for research reports that can be used by both the researcher and interested parties.

ORCID

Gamze Sarıkaş, https://orcid.org/0000-0001-6591-467X
Safiye Bilican Demir https://orcid.org/0000-0001-9564-9029

5. REFERENCES

Ağca, Ö. (2014, 25-27 Eylül). *Türkiye'deki üniversitelerin lisans ve lisansüstü programlarının özel eğitim açısından incelenmesi* [Conference presentation abstract]. 24. Ulusal Özel Eğitim Kongresi, Edirne, Türkiye.

Alper, A., & Gülbahar, Y. (2009). Trends and issues in educational technologies: A review of recent research in TOJET. *The Turkish Online Journal of Educational Technology - TOJET, 8*(2), 124-135.

Arık, R. S., & Türkmen, M. (2009). Eğitim bilimleri alanında yayınlanan bilimsel dergilerde yer alan makalelerin incelenmesi. Retrieved from http://www.eab.org.tr/eab/2009/pdf/488.pdf

Aydın, A., & Uysal, Ş. (2011). Türkiye'de ve yurt dışında eğitim yönetimi alanında yapılan doktora tezlerinin konu, yöntem ve sonuçlar açısından değerlendirilmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 14*(1), 177-201.

Başol, G., & Akın, U. (2006, 13-15 Eylül). *2001-2006 yılları arasında Türkiye'de eğitim alanında belli başlı indeksli dergilerde yayımlanan araştırma makalelerinin metodolojik bakımdan değerlendirilmesi* [Conference presentation abstract]. XV. Eğitim Bilimleri Kongresi, Muğla, Türkiye.

Benligiray, S. (2009). Türkiye'de insan kaynakları yönetimi alanında yapılan lisansüstü tezler ve bu tezlerde incelenen temaların analizi: 1983-2008 dönemi [The theme analysis of the

postgraduate theses written on human resource management in Turkey:1983-2008 period]. *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, *4*(1), 167-197.

Bilgin, N. (2006). *Sosyal bilimlerde içerik analizi -Teknikler ve örnek çalışmalar*. Ankara: Siyasal Kitapevi.

Büyüköztürk, Ş., & Kutlu, Ö. (2006). Sosyal bilim araştırmalarında yöntem sorunu. *Sosyal Bilimlerde Süreli Yayıncılık– 2006-I. Ulusal Kurultay Bildirileri* (pp. 113–122). Ankara: TÜBİTAK Yayını.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö.E., Karadeniz, Ş., & Demirel, F. (2016). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi Yayınları.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). New York: Routledge.

Coşkun, İ., Dündar, Ş., & Parlak, C. (2014). Türkiye'de özel eğitim alanında yapılmış lisansüstü tezlerin çeşitli değişkenler açısından incelenmesi (2008-2013) [The analysis of the postgraduate thesis written on special education in terms of various criteria in Turkey (2008-2013)]. *Ege Eğitim Dergisi*, *15*(2), 375-396.

Çiltaş, A., Güler, G., & Sözbilir, M. (2012). Türkiye'de matematik eğitimi araştırmaları: Bir içerik analizi çalışması. [Mathematics education research in Turkey: A content analysis study] *Kuram ve Uygulamada Eğitim Bilimleri*, *12*(1), 565-580.

Demirok, M.S., Bağlama, B., & Beşgül, M. (2015). A content analysis of the studies in special education area. *Procedia - Social and Behavioral Sciences*, *197*, 2459-2467.

Diken, İ.H., Ünlü, E., & Karaaslan, Ö. (2008). *Zihinsel yetersizlik ve yaygın gelişimsel bozukluk alanlarında lisansüstü tez bibliyografyası*. Ankara: Maya.

Diken, İ. H., Görgün, B., Öğülmüş, K., Kurnaz, E., & Baki, K. (2016). *Zihin yetersizliği ve otizm spektrum bozukluğu alanlarında lisansüstü tez bibliyografisi 2008-2015*. Ankara: Eğiten.

Doğan, H., & Tok, T. N. (2018). Türkiye'de eğitim bilimleri alanında yayınlanan makalelerin incelenmesi: Eğitim ve Bilim Dergisi örneği. *Curr Res Educ*, *4*(2), 94-109.

Doğru, M., Gençosman, T., Ataalkın, N.A. & Şeker, F. (2012). Fen bilimleri eğitiminde çalışılan yüksek lisans ve doktora tezlerinin analizi [Analysis of the postgraduate and doctoral theses conducted on sciences education]. *Türk Fen Eğitimi Dergisi*, *9*(1), 49-62.

Doğru, S.Y., Özlü, Ö., Kançeşme, C., & Doğru, S. (2015). Özel eğitim üzerine yapılan proje çalışmalarının değerlendirilmesi [Quantification of projects done on special education]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *15*(Özel Sayı), 286-296.

Erdem, D. (2011). Türkiye'de 2005–2006 yılları arasında yayımlanan eğitim bilimleri dergilerindeki makalelerin bazı özellikler açısından incelenmesi: Betimsel bir analiz. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *2*(1), 140-147.

Evrekli, E., İnel, D., Deniş, H., & Balım, A.G. (2011). Fen eğitimi alanındaki lisansüstü tezlerde yöntemsel ve istatistiksel sorunlar [Methodological and statistical problems in graduate theses in the field of science education]. *Ilkogretim Online - Elementary Education Online, 10*(1)*, 206-218.*

Göktaş, Y., Hasançebi, F., Varışoğlu, B., Akçay, A., Bayrak, N., Baran, M., & Sözbilir, M. (2012). Türkiye'deki eğitim araştırmalarında eğilimler: Bir içerik analizi. *Kuram ve Uygulamada Eğitim Bilimleri*, *12*(1), 443-460.

Güner-Yıldız, N., Melekoğlu, M.A., & Paftalı, A.T. (2016). Türkiye'de özel eğitim araştırmalarının incelenmesi [Special education research in Turkey]. *Ilkogretim Online - Elementary Education Online*, *15*(4), 1076-1089. https://doi.org/10.17051/io.2016.06677

Horst, P. (1966). *Psychological measurement and prediction*. Belmont: Wadsworth Pub. Co.

Karadağ, E. (2009). A thematic analysis on doctoral dissertations made in the area of education sciences. *Journal of Kırşehir Education Faculty*, *10*(3), 75-87.

Karasar, N. (2016). *Bilimsel araştırma ve yöntemi*. Ankara: Nobel Yayın Dağıtım.

Köklü, N., & Büyüköztürk, Ş. (1999). Eğitim bilimleri alanında öğrenim gören lisans-üstü öğrencilerin araştırma yeterlikleri konusunda öğretim üyelerinin görüşleri [Advisors' opinions about the research competencies of masters and doctorate students in educational sciences]. *Eğitim ve Bilim*, *23*(112), 18-28.

Kula-Kartal, S., & Mor-Dirlik, E. (2016). Geçerlik kavramının tarihsel gelişimi ve güvenirlikte en çok tercih edilen yöntem: Cronbach Alfa katsayısı [Historical development of the concept of validity and the most preferred technique of reliability: Cronbach alpha coefficient]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *16*(4), 1865-1879.

Kutluca, T., & Demirkol, M. (2016). Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi dergisinin bibliyometrik analizi [Bibliometric analysis of journal of Dicle University Ziya Gökalp Faculty of Education]. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, *28*, 108-118. https://doi.org/10.14582/DUZGEF.674

Kutluca, T., Birgin, O., & Gündüz, S. (2018). Türk Bilgisayar ve Matematik Eğitimi Dergisi'nde yayımlanmış makalelerin içerik analizi bağlamında değerlendirilmesi [Evaluation of the published articles in turkish journal of computer and mathematics education according to content analysis]. *Türk Bilgisayar ve Matematik Eğitimi Dergisi*, *9*(2), 390- 412. https://doi.org/10.16949/turkbilmat.332518

Lee, M.H., Wu, Y.T., & Tsai, C.C. (2009). Research trends in science education from 2003 to 2007: A content analysis of publications in selected journals. *International Journal of Science Education*, *31*(15), 1999-2020.

Melekoğlu, M.A. (2014). Special education today in Turkey. A. F. Rotatori, J. P. Bakken, S. Burkhardt, F.E. Obiakor, & U. Sharma (Eds.). *Special education international perspectives: Practices across the globe* (pp. 529-557). Bingley, UK: Emerald Group Publishing.

Metin, N.E. (2012). Özel Gereksinimli çocuklar. In E. Nilgün Metin (Ed.), *Özel gereksinimli çocuklar 1.* Ankara: Maya Akademi.

Ozan, C., & Köse, E. (2014). Eğitim programları ve öğretim alanındaki araştırma eğilimleri [Research trends in curriculum and instruction]. *Sakarya University Journal of Education*, *4*(1), 116-136.

Özenç, E.G., & Özenç, M. (2013). Türkiye'de üstün yetenekli öğrencilerle ilgili yapılan lisansüstü eğitim tezlerinin çok boyutlu olarak incelenmesi [The multidimensional examination of master-doctorial dissertations made in Turkey about gifted and talented students]. *Türkiye Sosyal Araştırmalar Dergisi, 171*, 13-28.

Seçer, D., Ay, D., Ozan, C., & Yılmaz, B.Y. (2014). Rehberlik ve psikolojik danışma alanındaki araştırma eğilimleri: Bir içerik analizi [Research trends in the field of guidance and psychological counseling: A content analysis]. *Turkish Psychological Counseling and Guidance Journal*, *5*(41), 49-60.

Staton, A.Q.S., & Wulff, D.H. (1984). Research in communication and instruction: Categorization and synthesis. *Communication Education*, *33*(4), 377-391.

Şenyurt, S., Özer-Özkan, Y. (2017). Eğitimde ölçme ve değerlendirme alanında yapılan yüksek lisans tezlerinin tematik ve metodolojik açıdan incelenmesi [A thematic and methodological analysis of master's dissertations in the field of measurement and evaluation in education]. *Ilkogretim Online - Elementary Education Online, 16(2), 628-653,* 2017. https://doi.org/10.17051/ilkonline.2017.304724

Ünlü, Ö., Güner-Yıldız, N., & Aktar, E. (2020). Investigation of doctoral dissertations in special education in Turkey. *Ilkogretim Online - Elementary Education Online, 19*(2), 923-931. https://doi.org/10.17051/ilkonline.2020.695825

Tavşancıl, E., Çokluk, Ö., Gözen Çıtak, G., Kezer, F., Yalçın Yıldırım, Ö., Bilican, S., Büyükturan, B.E., Şekercioğlu, G., Yalçın, N., Erdem, D., & Özmen, T.D., (2010). *The*

investigation of theses completed at the institutes of educational sciences (2000-2008). Ankara University Scientific Research Project Final Report, Ankara.

Tiryaki, E.N. (2017). Evaluation of the postgraduate theses written in the field of special education in terms of language education and teaching. *International Online Journal of Educational Sciences, 9(2),* 454-463.

Tiryakioğlu, Ö. (2014). Content analysis of the articles published in the Ankara University Special Education Journal within the years 2004-2013. *Procedia - Social and Behavioral Sciences, 143*, 1164-1170.

Tosuntaş Ş.B., Emirtekin E., & Süral İ., (2019). Eğitim ve öğretim teknolojileri konusunda yapılan tezlerin incelenmesi (2013-2018) [Examination of theses on educational and instructional technologies (2013-2018)]. *Journal of Higher Education and Science*, *9*(2), 277-286. https://doi.org/10.5961/jhes.2019.330

Turan, S., Karadağ, E., Bektaş, F., &Yalçın, M. (2014). Türkiye'de eğitim yönetiminde bilgi üretimi: Kuram ve uygulamada eğitim yönetimi dergisi 2003-2013 yayınlarının incelenmesi [Knowledge production in educational administration in Turkey: An overview of researches in journal of educational administration: Theory and practice - 2003 to 2013-]. *Kuram ve Uygulamada Eğitim Yönetimi*, *20*(1), 93-119. https://doi.org/10.14527/kuey.2014.005

Varışlıoğlu, B., Şahin, A., & Göktaş, Y. (2013). Türkçe eğitimi araştırmalarında eğilimler [Trends in turkish education studies]. *Kuram ve Uygulamada Eğitim Bilimleri*, *13*(3), 1767-1781. https://doi.org/10.12738/estp.2013.3.1609

Yalçın, S. (2016). Ölçme ve değerlendirme alanındaki dergilerde yayımlanan makalelerin içerik analizi [Content analysis of research articles in measurement and evaluation journals]. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, *49*(1), 65-84. https://doi.org/10.1501/Egifak_0000001375

Yalçın, S., Yavuz, H.Ç., & İlgün Dibek, M. (2016). A review of articles published in educational journals having highest impact factors: Content analysis. *Eğitim ve Bilim*, *40*(182), 1-28. https://doi.org/10.15390/EB.2015.4868

Yaşar, Ş., & Papatğa, E. (2015). İlkokul matematik derslerine yönelik yapılan lisansüstü tezlerin incelenmesi [The analysis of the graduate theses related to mathematics courses]. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, *5*(2), 113-124.

Yılmaz, K., & Altınkurt, Y. (2012). An examination of articles published on preschool education in Turkey. *Kuram ve Uygulamada Eğitim Bilimleri*, *12*(4), 3227–3241.