



Confusion of Scale Development: Investigation of Self-Efficacy Scales

Ölçek Geliştirme Karmaşası: Öz-yeterlik Ölçeklerinin İncelenmesi

Fulya BARIŞ PEKMEZCİ* Cansu AYAN**

• Geliş Tarihi: 20.02.2019 • Kabul Tarihi: 16.05.2019 • Çevrimiçi Yayın Tarihi: 21.06.2019

Abstract

The aim of this study is to determine to which extent the measurement tool development steps that should be followed during the development of self-efficacy scales used in education and psychology are met. In line with this general purpose, the following questions tried to be answered for each study that was examined regarding the five sections, which were introduction, theoretical background, item writing, validity and reliability (for first and second implementation). In the scope of the study, document analysis was implemented. The self-efficacy development studies in the national and international literature between 1997 and 2018, which were possible to get in the electronic environment, were examined in the scope of the research. It was determined that one of the deficiencies often seen when the self-efficacy scales are examined is that the psychological structure is not defined operationally. And also the mistake that was often encountered in scale validity studies was seen to be making both explanatory and confirmatory factor analyses on the same data set. One of the most remarkable mistakes done is the confusion of scale and questionnaire. That there is a confusion regarding concepts means the scale developers do not know the difference between a scale and a questionnaire yet.

Keywords: self-efficacy, self-efficacy scale, APA standards for educational and psychological testing, scale development procedure

Cited:

Bariş Pekmezci, F., & Ayan, C. (2020). Confusion of scale development: Investigation of self-efficacy scales. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 48, 130-151. doi: 10.9779/pauefd.529986

* Ass.Prof.Dr., Yozgat Bozok University, Yozgat, fulyabaris@gmail.com, Orcid ID: <https://orcid.org/0000-0001-6989-512X>.

** Res. Assist. Dr., Ankara University, Ankara, cnsayan@gmail.com, Orcid ID: <https://orcid.org/0000-0002-0773-5486>.

Öz

Bu araştırmanın genel amacı, eğitim ve psikolojide kullanılan öz-yeterlik ölçeklerinin geliştirilmesi aşamasında izlenmesi gereken ölçme aracı geliştirme adımlarının ne derece karşılandığını ortaya çıkarmaktır. Bu genel amaç doğrultusunda giriş bilgileri, kuramsal kısım, madde yazımı, geçerlik ve güvenirlik ve ikinci uygulama olmak beş bölüme ilgili incelenen her çalışma için ilgili sorulara yanıt aranmıştır. Çalışma kapsamında doküman analizi yapılmıştır. Elektronik ortamdan ulaşılabilen 1997-2018 yılları arasında yerli ve yabancı literatürde yer alan öz-yeterlik ölçeği geliştirme çalışmaları incelemeye alınmıştır. Araştırmaya dahil edilecek çalışmalar seçilirken ölçülmesi hedeflenen öz-yeterlik kavramının eğitim ve psikoloji ile ilişkili olmasına, ve ölçeğin tüm maddelerinin ulaşılabılır olmasına dikkat edilmiş, bu kriterlere uymayan çalışmalar araştırmaya dahil edilmemiştir. Öz-yeterlik ölçekleri incelendiğinde sıklıkla görülen eksikliklerden biri psikolojik yapının işevuruk tanımının yapılmaması olduğu görülmüştür. Ayrıca ölçek geçerliği çalışmalarında sıklıkla karşılaşılan hatanın aynı veri seti üzerinden AFA ve DFA yapılması olduğu görülmüştür. Yapılan hatalardan en göze çarpanlardan bir diğerinin ise ölçek-anket kavram karmaşasının yaşanmış olmasıdır. Kavram karmaşasının olması ölçeği geliştirenlerin henüz ölçek ve anket ayrımını bilmedikleri anlamına gelmektedir.

Anahtar sözcükler: öz-yeterlik, öz-yeterlik ölçeği, ölçek geliştirme, APA eğitimde ve psikolojide ölçme standartları, ölçek geliştirme aşamaları

Atıf:

Barış Pekmezci, F., ve Ayan, C. (2020). Ölçek geliştirme karmaşası: Öz-yeterlik ölçeklerinin incelenmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 48, 130-151. doi: 10.9779/pauefd.529986

Introduction

Affective characteristics are not directly observed and measured by its nature. For this reason, indirect measurement methods are being used in order to measure the affective characteristics. The most commonly used indirect measurement method is implementing psychological tests (Anastasi, 1988). These implementations are done by scoring the reactions an individual shows towards the items in the test/scale according to a method. Psychological tests that are often preferred are easier to implement than the other data collection methods such as observations and interviews. That psychological tests can be scored objectively, create a chance to obtain data from large groups in one time and provide an opportunity to make a valid and reliable observation are among the biggest advantages they have (Conway, 2006; Cronbach, 1990). The scales, which are one of the psychological tests, are among the commonly used measurement tools to collect data in terms of a target individual or individuals, topic or content besides showing the mathematical qualities of the measurement results basically (Yurdugül, 2005).

When the literature is reviewed, it is seen that the researchers either use the existing scales or develop their own ones. There can be two cases that push researchers to develop a new scale. One of them is that there has not been a tool developed to serve the researcher's purpose; the other one is that the existing scales have been outdated or do not possess the features needed. These two cases make it inevitable for the researchers to develop a new scale. But it is not right to develop a scale without knowing the notion that is going to be measured or without knowing the measurement process despite knowing the notion to be measured (Erkuş, 2012). If there are any doubts regarding the variable's measures, the result obtained via the relation and difference tests, which are analysed, based on these suspicious measures and the interpretations that are made upon these are going to be false (Crocker and Algina, 1986). This is because it is not going to be possible to exhibit the structure with the findings obtained through the aforementioned scale. As a result of this, the measurements will not exhibit the true values of the related structure. In other words, there will be a measurement error.

It is quite important that the scores obtained via the psychological tests are free from errors for the reliability and validity of the measurement results. To provide this is possible only by following some standards. To take these standards that have been developed to prevent the possible measurement errors in the test development process into consideration provides an opportunity to have more quality measurement tools. In this context, the purpose of the standards is to provide scales for the test implementations to be developed and evaluated and to be a guide to assess the validity of the test implementation and the interpretations made upon the test scores (APA, 2014). Errors can be reduced to the minimum with the test development processes to be followed that is made by taking these standards into consideration. According to Crocker and Algina (2008) and DeVellis (2003), the standard phases to be followed while developing measurement tools are as follows:

- to determine the structure to be measured
- to determine the purpose of the scale
- to determine the theoretical bases and the pragmatic definitions
- to create an item pool
- to determine the type of the scale (Thurstone, Likert, Guttman, etc.)

- to ask for experts' opinions on the first item pool created
- to determine the items to be in the scale after the experts' opinions
- to have a pilot implementation on the study group
- to assess and evaluate the items (item analysis, validity, reliability analyses, factor analysis etc.)
- to finalize the scale
- to implement for the second time (Confirmatory Factor Analysis)

Researchers often develop scales in the fields of education and psychology. In the literature review done, it was seen that the reliability and validity studies were generally done regarding the scales that were in the development phase. However, it was also seen that no revisions of most scales were done against the possibility of being outdated in the following years. In this context, it is necessary to check the quality of aforementioned scales. For this purpose, there is only one study done in 1971 by the Ministry of National Education (MoNE) Administration of Planning, Research and Coordination Department of Guidance and Psychological Counseling which is about examining the qualities of the scales that have been developed all over the country. In addition, various national and international studies in which the psychological tests are examined in terms of their psychometric qualities have been conducted (Acar Güvendir and Özer Özkan, 2015; Çüm and Koç, 2013; Delice and Ergene, 2015; Doğan 2009; Erkuş, 2016; Hinkin, 1995; Gül and Sözbilir, 2015; Mor Dirlik, 2014; Slavec and Drnovsek, 2012; Tavşancıl, Güler and Ayan, 2014; Tosun and Taşkesenligil, 2015; Özge, 1981; Worthington and Whittaker, 2006). The psychological tests were taken as a framework in almost all these studies and there were not made any examinations regarding the qualities of the items in the scale that were written. According to Anastasi (1988), the validity and the reliability of a measurement tool are directly related to the characteristics of items in the scale. In line with this, while examining a scale, whether the scale has some features of the structure that it aims to measure should also be taken into consideration. As a result, findings regarding the validity of the tool will be able to be obtained. For this reason, examining all scales specifically rather than doing this from a general perspective will bring out more functional results in evaluating the current scales. In other words, knowing the phases to develop a scale is necessary but not a sufficient condition to develop a quality scale. It is also important to have a full knowledge on the theoretical background that is measured. In the studies conducted in the fields of education and psychology the affective features of the students are often discussed and depending on this there have been studies on scale development. One of the aforementioned features is "self-efficacy". Self-efficacy is an important concept, which Bandura (1977) thinks as effective on behaviours and is emphasized in Social Learning Theory. The belief of an individual in possessing the skill, the attitude and the information needed to be able to do a work or implementation is defined as "self-fficacy" (Bandura, 1994). According to Zusho and Pintrich (2003), self-efficacy is the attitudes of individuals regarding their capacities in performing a task or their beliefs in themselves in doing a job. Self-efficacy is a concept that does not mean that the individual is talented but he believes in his own sources. Bandura (1997) states that as the perceived self-efficacy gets stronger, the determination in realizing a target gets stronger too. An individual who has sufficient skills to deal with a problem but has a low level of self-efficacy will not be able to put these skills into an action. The concept of "self-fficacy" consists of planning an action, being aware of necessary skills and organizing them,

the motivation level that comes out as a result of going through the outcomes achieved with difficulties (Bandura, 2006).

When examined specifically, the perceived self-efficacy does not have a measure that serves every purpose. The descriptor and predictor value of the approach that serves everything is usually limited (Bandura, 2006). This case leads to ambiguities about what is measured exactly. The perceived self-efficacy scales should be configured according to the special area of the operation with the object/feature of interest (Bandura, 2006; Pajares, 1996; Zimmerman, 1995). Namely, self-efficacy scales should be measuring the competence regarding the content/topic in contrast to emphasizing a general self-efficacy.

Self-efficacy, by its nature, includes an individual's self-belief in an object or a feature. Therefore, while writing items for the self-efficacy scale, attention must be paid that the items are able to reflect the structure very well. Bandura (2006) states that the expressions used in developing the self-efficacy scale should be formed with "can". In addition, he emphasizes that to grade the items by 10 intervals between 0 and 100 is to be paid attention in terms of the measurement precision. In line with this, he also states that the item that includes only a few reply categories are less sensitive and less reliable. Moreover, Bandura (2006) emphasized that competence scales are unipolar and they do not consist of negative numbers because the judgments regarding 'not being competent' (0) can not be less than zero (0). The negative grades below (0) in unipolar scales are not significant in a certain competence level.

It is possible to encounter studies on developing self-efficacy scales in both national and international journals. While developing a self-efficacy scale, it is expected to take the steps of developing a measurement tool into consideration like other scales. However, when the literature is reviewed, there have been encountered many scales that were developed without following the steps of developing a standard test/measurement tool. In addition, it is observed that there are items, which do not include expressions that the feature measured (self-efficacy) requires in most of these scales that were developed. Using the scales that were developed in this way and interpreting the scores obtained from these scales statistically cause misleading the science. Each day a new one is added to this stack, which consists of the scales that cause creating a false basis for the future researches. For this reason, the existence of a high number of self-efficacy scales that were developed but impossible to use is a sign of a big effort and time loss. The validity and reliability of the results obtained from the studies in which these scales were used must be questioned. There have been encountered studies including the general investigations that belong to the scales developed in the fields of education and psychology. There have not been encountered any studies in which specifically self-efficacy scales are examined. Every psychological structure is unique and therefore should be examined according to its frame. Hence, the idea that putting forward the problems during the development of self-efficacy scales is an important step to solve these problems creates a necessity for this research to be done. This research is seen as important in terms of describing the existing problems in the literature and being informative for the future studies on developing self-efficacy scales.

The overall purpose of this study is to determine to which extent the measurement tool development steps that should be followed during the development of self-efficacy scales used in education and psychology are met. In line with this general purpose, the following questions were tried to be answered for each study that was examined regarding the five sections, which

were introduction, theoretical background, item writing, validity and reliability, and the second implementation.

1. To determine the purpose of the scale
 - 1.1. Are the reasons why a new scale is needed explained?
 - 1.2. Is the purpose of the scale stated sufficiently?
 - 1.3. Is the target population of the scale stated clearly?
2. Theoretical Background and Pragmatic Definitions
 - 2.1. Is the theoretical background of the psychological structure that is to be measured (self-efficacy) stated?
 - 2.2. Is the measurement tool based on a measurement theory?
 - 2.3. Is the structure that is to be measured defined operationally?
3. Item Writing and Pre-testing
 - 3.1. Item Writing
 - 3.1.1. Is there any information regarding the item writing process?
 - 3.1.2. Are the items identifiable for the target population by everyone in the same way?
 - 3.1.3. Are the unnecessary words avoided when writing items?
 - 3.1.4. Is each item capable of measuring a single topic?
 - 3.1.5. Are the grading expressions compatible with the item structures?
 - 3.1.6. Are there any factual expressions in the articles?
 - 3.1.7. Is there any expert opinion regarding the items prepared?
 - 3.2. Pre-Testing
 - 3.2.1. Is the group of the pre-test similar to the target population?
 - 3.2.2. Is the sample size large enough?
4. Validity- Reliability Studies
 - 4.1. Are the validity studies conducted?
 - 4.2. Are the reliability studies conducted?
 - 4.3. Are the validity studies suitable?
 - 4.4. Are the reliability studies suitable?
5. Second Implementation
 - 5.1. Is the revised version of the scale implemented?
 - 5.2. Is the statistical information regarding this implementation included?

Data Collection Tools and Techniques

The encoding list was taken from Tavşancıl, Güler and Ayan (2014) which was developed for attitude scales. This encoding list revised based on the steps to consider while developing a self-efficacy scale and the purpose of the research. This list was revised from The opinions of three Measurement and Evaluation experts were taken on the encoding list that was developed. The last version of the encoding list was formed as a result of the revisions of the experts.

In the first part of the encoding list is the information related to the publication date of the article and the journal in which it was published. There is a control list, which consists of five parts and has expressions graded as “yes”, “partially” and “no” in the second part. In the final part, on the other hand, are the open-ended questions, which are there to obtain more detailed information, regarding some points in this encoding list. In addition, inappropriate item samples are presented after a detailed examination of the items in the scale in the last part of the encoding list.

Data Analysis

The research data were subjected to the content analysis, which is a type of analysis used in qualitative researches. The categorical analysis method, which is a kind of content analysis, was applied and frequencies of each category were calculated (Tavşancıl and Aslan, 2001).

The coding process was carried out by two encoders, and it was definitively agreed on what was expected in each item and what was to be marked. All three studies were coded altogether by the encoders, and all studies were also coded separately by both encoders. As a proof for the reliability, the consistency between the codings done by two different researchers was checked by calculating the reliability coefficient between the encoders.

The formula that Miles and Huberman (1994) suggested and is generally used in determining the reliability of the content analysis studies was used in order to calculate the percentages of fit between the researchers. It is expected that the percentage of fit between the markers is above 70% (as cited in Tavşancıl and Aslan, 2001). According to the formula that is “Reliability= number of agreement / (number of agreement + number of disagreement), the reliability coefficient between the encoders was calculated as 0.89, the parts where the fit was not met were checked again and a common decision was made.

Results

Determine the Purpose of the Scale

In line with this sub-purpose, three questions asking for the reasons for the need for a new scale are, the purpose of the scale and whether the target population was announced were tried to be answered. In this context, if the relevant questions were all answered, it was marked as “yes”; if there were some missing parts, it was marked as “partially” and if there was no information, it was marked as “no”. The frequency values obtained are presented in Table 2 below.

Table 2. Frequency Values Regarding the Determination of the Purpose of the Scale

	National			International		
	Y	P	N	Y	P	N
Determining the purpose of the scale						
Are the reasons for the need for a new scale explained?	10	10	8	18	8	-
Is the purpose of the scale stated sufficiently?	20	7	1	20	6	-
Is the target population of the scale stated clearly?	20	5	3	24	2	-
Total	50	22	12	62	16	0

When Table 2 is examined, in the international literature there are 16 studies that have the relevant information partially, leave some ambiguities and do not present the whole information needed. It is seen that there are 22 studies that have missing information and 12 studies that have no relevant information in the national literature. Based on this result, it is possible to say that the problem in this issue is mostly seen in the national literature in the scope of the studies examined. Regarding the reasons for the need for a new scale, information about the reasons for the scale to be developed should be presented at the beginning of the research reports. In addition, it was also asked to see whether there were any similar scales in the relevant literature in the studies. However, there was not any information encountered in some of the studies. This case brings the conclusion that the literature is insufficient in order to determine the need for the scale to be developed and to specify the necessity. This is thought to be possible to cause a big effort and time loss. As to the determination of the target population, there have been specified some problems in giving the exact definition of the target population. For example, the scale that was developed was said to be aimed at the students however no information was presented regarding which class level of the students this scale would be implemented. When the items of the scale, on the other hand, were examined, it was concluded that it would not be suitable for every class level. Moreover, that the pilot testing was implemented only on the students in the 6th, 7th and 8th grades called into doubt that there was an age limitation but it was left blank while filling in the report. This may lead to problems for the future usages of the scale. For this reason, the researcher should clearly state the target population of the scale that he developed

Theoretical Background and Operational Definitions

In line with this sub-purpose, the theoretical background of the psychological structure that was demanded to be measured, the measurement theory that the scale was based on and the pragmatic definition of the psychological structure were questioned. This was done via three questions. The relevant questions and the frequency values of the questions are presented in Table 3.

Table 3. Frequency Values About Theoretical Background

	NATIONAL			INTERNATIONAL		
	Y	P	N	Y	P	N
Theoretical Background and Pragmatic Definitions						
Is the theoretical background of the psychological structure to be measured stated?	8	20	0	18	6	2
Is the measurement tool based on a measurement theory?	-	-	28	-	2	24
Is the psychological structure to be measured defined operationally?	5	5	18	10	10	6
Total	13	25	46	28	18	32

When the Table 3 is examined, it is seen that the theoretical background of the psychological structure to be measured is reported to a large extent even if it is partially done. Both national and international studies lack information as to providing pragmatic definitions. In the scope of the scale development studies, a measurement theory should also be present besides the theoretical background of the psychological structure. It is seen that not only this theory stays hidden but also no justification and a conscious preference are made while choosing a measurement theory in the frame of the studies conducted in the fields of education and psychology. Since this was thought to be a deficiency for the field and was wished to be highlighted, this item was added to the form. The results also supported the predictions. For instance, a study in the international literature stated that it adopted the item response theory as the measurement theory and conducted the analyses in line with that. However, the researchers thought it was suitable to mark this study as ‘partially’ since it was thought to lack in providing enough justification. On the other hand, when the results of the detailed reading that is done in the scope of the mentioned sub-purpose are examined, the most common mistake is having a mechanical logic while presenting the theoretical background. As a result of this, the link between the scale the researchers developed and the theoretical background that they presented in their researches was not seen. For instance, the number of the studies that made use of the theoretical backgrounds in determining and naming the sub-dimensions of the scale is so little, and as a result of this it was seen that the sub-dimensions and their names were mostly irrational.

Item Writing and Pre-Testing

In line with this sub-purpose, presenting the information regarding the item writing process, taking some experts’ opinions on the items prepared and the properties that the items should possess were questioned. A total of seven questions that question these cases were tried to be answered. These questions and the relevant frequency values are presented in Table 4.

Table 4. Frequencies Regarding the Items and Item Writing Process

Item Writing and Pre-test	NATIONAL			INTERNATIONAL		
	Y	P	N	Y	P	N
Is there any information regarding the item writing process?	17	8	3	22	4	-
Are the items identifiable for the target population by everyone in the same way?	21	6	1	22	3	1
Are the unnecessary words avoided when writing items?	16	9	3	21	1	4
Is each item capable of measuring a single topic?	19	7	2	22	3	1
Are the grading expressions compatible with the item structures?	21	3	4	26	-	-
Are there any factual expressions in the articles?	5	4	19	5	3	18
Is there any expert opinion regarding the items prepared?	-	24	4	4	13	9
Total	99	43	36	122	27	33

When Table 4 was examined, the international articles had almost no problems in terms of the information provided about item writing process, the intelligibility of the items and grading expressions when compared to the national articles. On the other hand, there were seen mistakes for every subject in the national articles. According to the information presented regarding item writing process, it was detected that researchers often made use of the items in

other scales present in the literature and secondly they formed the item pools after making essays written. As to taking expert opinion on the items prepared, the frequency of ‘partially’ option is seen to be quite high. Here, it was thought that the scale developers should consult at least three experts who were a field expert, a language expert and a testing and evaluation expert on their opinions for the items and only the studies in which these three experts were consulted were marked as “yes”. When the detailed notes taken about which experts’ opinions were taken were examined, the tendency was seen to be consulting a field expert. The number of the studies in which the opinions of a language expert and a measurement and evaluation expert were taken was detected to be quite low whereas developing a scale is one of the professions of the measurement and evaluation experts. If they are included in the process by having this consciousness, the mistakes made are thought to reduce.

Improper item samples were also determined besides the frequencies of the mistakes done regarding the structure of the items. The sample items are presented below:

National Literature

- Okuduğum metnin yazarı hakkında bilgi edinirim / I get information about the author of the text that I read.
- Metni okumadan önce metnin içeriği hakkında çeşitli sorular hazırlarım / I prepare various questions about the content of the text before I read it.
- Dans çalışmalarına katılmaya istekliyim / I am willing to participate in dance activities) (*expression of attitude*)
- Yeni bir hareket öğrenimine açığım / I am open to learning a new move. (*expression of attitude*)
- Dans ederken kendimi özgür hissederim / I feel free while dancing.
- Grup içinde dans etmekten keyif alırım / I enjoy dancing in groups.
- Yeni danslar öğrenmekten mutlu olurum / I am happy to learn new dance types. (*expression of attitude*)
- Piyano dersine her dönemde aynı ilgiyi gösterebileceğime inanıyorum / I believe that I can have the same interest in piano lessons in each term. (*expression of interest*)
- Piyano dersinde öğrendiklerimin başka derslerde de faydalı olduğuna inanıyorum / I believe that what I learn in the piano lesson is also useful for other lessons.
- Piyano dersinde öğrendiklerimin meslek hayatımda bana yardımcı olacağına inanıyorum / I believe that what I learn in the piano lesson will be helpful for me in business life. (*expression of care*)
- Öğrencilerin Türkiye’deki çevre problemlerini öğrenmesinden sınıf öğretmeni sorumludur / The elementary school teacher is responsible for the students’ learning the environmental problems in Turkey. (*expression of opinion*)
- Anlamadığım bir çevre konusunu öğrenmek için kimlerden yardım alabileceğimi bilmiyorum / I don’t know who I can get help from to learn about an environmental issue that I do not understand.

- Matematik ve Türkçe derslerinde ders planlarını, sözgelimi I. Devre olarak adlandırılan A grubunda bir sınıfın “öğretmenli” olduğu bir ders saatinde “ödevli” olan diğer iki sınıfı, seviye grupları temelinde çalışmalarını sağlayacak biçimde biçimlendirmede kendimi yeterli görüyorum / I consider myself to be successful in shaping lesson plans in Mathematics and Turkish lessons. For instance, in the class hour when a class is “with their teacher” in group A called as the 1st.period, I can arrange other two classes “with their assignments” to study on the basis of their levels. *(It is not clear, too long, not being understood)*
- Geometri ile el becerilerimi arttırabileceğimi düşünüyorum / I think I can improve my hand skills with geometry.
- Geometri sorusu çözdükçe kendime olan güvenimin artacağını düşünüyorum / As I solve a problem in geometry, I believe that my self-confidence will increase.
- Başarısı düşük olan öğrencilerin matematikte ilerlemesi öğretmenin daha fazla ilgi göstermesinden kaynaklanır / The development of students with low success in Math stems from the teacher’s showing more interest. *(factual expression)*
- Günlük yaşamda kendimi İngilizce yazılı olarak ifade edebilirim / I can express myself in English in everyday life. *(incomprehensibility)*
- Bir öğrenci herhangi bir müzik kavramını öğrenme konusunda güçlük çekiyorsa, o öğrencinin o kavramı daha iyi anlamasına nasıl yardımcı olabileceğimi bilemiyorum / If a student has difficulty in learning a term in music, I do not know how to help that student learn that term better. *(unnecessary words)*
- Okuduğum her metni anlayabilirim / I can understand every text that I read. *(a too general judgment)*
- Okuma etkinliklerinde sıkılmadan dikkatimi sürdürebilirim / I can keep my attention in reading activities without getting bored. *(incomprehensibility)*
- Eğer seçim şansım olsaydı Türkçe dersini almak istemezdim / I would not take Turkish lessons if I had a chance to choose *(expression of attitude)*
- Zor olan Türkçe konularını anlayabileceğimden çok emin değilim / I am not completely sure that I can understand the difficult topics in Turkish. *(expression of incomprehensibility and frequency)*
- Türkçe dersinde uzun cümleler ve paragrafların biçim ve anlam özellikleri beni endişelendirir / The long sentences and the form and the meaning properties of the paragraphs in Turkish lessons worry me. *(incomprehensibility)*
- Türkçe dersinde sözcük, cümle ve paragraf düz soruları çözerken zorlanırım / I have difficulty in solving questions regarding words, sentences and paragraphs in Turkish lessons. *(incomprehensibility, more than one judgement)*
- Türkçe dersinde anlatılanları dinlerken tam olarak anlatılanları anlayamıyorum / While listening to what is being told in Turkish lessons, I cannot understand it exactly. *(unnecessary words)*

International Literature

- Science and technology tests disconcert me. (Expression of attitude)
- If it was optional, I would not want to learn science and technology lesson. (Expression of attitude)
- I am expecting that, I will be successful in science and technology lesson. (Expression of expectation)
- If scientific and technologic activities are hard, I give them up or I accomplish parts, which are easier. (More than one case mentioned)
- I'm confident that I can choose recreational activities in which I am interested.
- I use the computer system as much as possible.
- I work hard in school.
- Most of my classmates like to do math because it is easy (not a personal judgment)
- I will graduate from high school. (a future prediction)
- I go to a good school. (An opinion regarding school)
- Sometimes I think an assignment is easy when the other kids in class think it is hard.
- Adults who have good jobs probably were good students when they were kids.
- When I am old enough, I will go to college. (A future plan).
- No one cares if I do well in school.
- What I learn in school is not important. (Expression of care)
- Understand when a colleague is irritated with you. (Social awareness)
- Understand the mood of your colleagues. (Social awareness)
- Put yourself in the shoes of a work colleague who is in trouble. (social awareness)

The questions of the similarity between the group and the target population and the size of the trial group were also tried to be answered. In order for the group size to be sufficient enough, the number of the participants was kept as ten times more than the number of items. (Kline, 1994). The relevant frequencies are presented in Table 5.

Table 5. Frequencies Regarding Pre-test

Item Writing and Pre-test	NATIONAL			INTERNATIONAL		
	Y	P	N	Y	P	N
Pre-test						
Is the trial group similar to the target population?	24	4	-	26	-	1
Is the sample size large enough?	22	-	6	24	-	2
Total	46	4	6	50	0	3

When Table 5 is examined, it seems to be only one article that has a problem regarding the pre-test in the international literature. When the national articles are checked, it is seen that a problem regarding the group's similarity actually stems from the fact that the target population is not defined well. The sample size is not convenient to work with in six of the national articles.

Validity and Reliability Studies

In line with this sub-purpose, the relevant frequencies in which whether validity and reliability studies are done and the convenience of these studies in terms of both statistical techniques and being reported are presented in Table 6.

Table 6. Frequencies Regarding Validity and Reliability Studies

4. Validity-Reliability Studies	NATIONAL			INTERNATIONAL		
	Y	P	N	Y	P	N
Are the validity studies conducted?	27	-	1	25	-	1
Are the reliability studies conducted?	27	1	-	25	1	-
Are the validity studies suitable?	17	10	1	17	8	1
Are the reliability studies suitable?	20	7	1	25	-	1
Total	91	18	3	92	9	3

When Table 6 is examined, it is seen that all articles, except for one, present validity and reliability proofs among both national and international articles in the scope of the articles examined. It is detected that in only one national article while the proof related to validity is presented; there seem no studies regarding the reliability. Additionally, when the convenience of these studies conducted is investigated, the general problem is seen to stem from making confirmatory factor analysis on the data set where explanatory factor analysis is made before.

Also it was seen that there were two studies in which items were eliminated during explanatory factor analysis but it was continued with confirmatory factor analysis without collecting new data afterwards. There was not done any factor analyses in one of the studies and frequency values were calculated instead of factor loadings.

Second Implementation

In line with this sub-purpose, relevant frequencies regarding the implementation of the final form of the scale and the calculation of the relevant statistics are presented in Table 7.

Table 7. Frequencies regarding the second implementation

5. Second Implementation	NATIONAL			INTERNATIONAL		
	E	K	H	E	K	H
Is the revised version of the scale implemented?	3	-	25	1	-	25
Is the statistical information regarding this implementation included?	2	1	25	1	-	25
Total	5	1	50	2	0	50

When the scale development studies in the literature were examined, it was seen that there were made some changes in the scale such as eliminating items after explanatory factor analysis. However, it was thought that there were made no studies as to making the final form of the scale. With the idea that this phase is an important and necessary step, some questions regarding this expectation were added to the checklist. However, as also seen in Table 6, the result supported the prediction, it was detected that no articles had such a study conducted except for three national and one international article. In one of the articles in which the second implementation was done, after the second implementation a t-test that showed whether there

was a difference between the girls and boys was conducted over the last data. However, due to the lack of the expected analyses, while it was marked as ‘yes’ for the second implementation, it was marked as ‘partially’ in terms of the question that was about the statistical knowledge.

Discussion and Results

The general purpose of this research is to put forward to what extent the standard steps of developing a measurement tool are followed in developing self-efficacy scales used in education and psychology. In line with this general purpose, self-efficacy scales that were developed between 1997 and 2018 were examined in five sections as introductory information, theoretical background, item writing, validity and reliability and second implementation.

One of the deficiencies often seen when the self-efficacy scales are examined is that the psychological structure is not defined operationally. All measurement processes start with determining the quality of the property to be measured. In line with the theoretical definition of the variable that is aimed to be measured, the variable needs to be defined pragmatically in an observable and measureable way (Tezbaşaran, 2008). All processes and the procedures after this phase are going to be shaped in accordance with the definition that is to be made (Erkuş, 2012). Also, while mentioning the theoretical background of the psychological structure in the scales developed, this background was not correlated to a measurement theory. The study in which Çüm and Koç (2013) examined the articles about scale development and implementation supports this finding. Almost all scales lack in the measurement theory that the researches are based on.

According to the information given about item writing process, it was detected that researchers often made use of the items in other scales present in the literature and secondly they formed the item pools after making essays written. As to taking experts’ opinion on the items prepared, the general tendency was seen to be consulting only a field expert. It was seen that the number of the studies in which testing and evaluation and language experts’ opinions were taken is quite low. These findings are supporting the studies of Çüm and Koç (2013) and Dirlik (2014).

During the investigations made in the scope of the research, the mistake that was often encountered in scale validity studies was seen to be making both explanatory and confirmatory factor analyses on the same data set. In addition, after eliminating items from the scale, the second explanatory factor analysis was also seen to be made on the same data set. The data set that is present after item elimination is the data set which still has the effect of the item eliminated. For this reason, new data need to be collected after eliminating items and explanatory factor analysis should be made on these new data. These mistakes indicate that researchers do not have enough information about factor analysis. The findings regarding factor analysis are consistent with the finding of the study, which belongs to Kaya Uyanık et al (2017).

One of the most remarkable mistakes done is the confusion of scale and questionnaire. That there is a confusion regarding concepts means the scale developers do not know the difference between a scale and a questionnaire yet. In another inaccurate scale detected, the scale was developed in Turkish but the article was translated into English due to the publishing

language policy of the journal. Translating the items in a scale developed in a language into another language is only possible with scale adjustment steps. Translating the scale into English and publishing it may lead to misunderstandings and misuses.

Scale development is an area that requires expertise. Therefore the researchers who are going to develop a scale should definitely get support from testing and evaluation experts in the process. If they are included in the process by having this consciousness, the mistakes made are thought to reduce. As a result of the investigations, major problems were detected in self-efficacy scale development studies. This problem may be seen as a sign of the fact that scale development steps are not understood well enough. Although the number of the scale development studies has increased in recent years, it is strange that their quality is still low. Even if scale development steps are followed appropriately, there have been encountered items misinterpreted in many self-efficacy scales examined. Scale development is a holistic process. This process requires mastering both the development steps and the theoretical background of the psychological structure. Without mastering the theoretical background of the psychological structure, the indicators of that structure are not going to be defined accurately and as a result there are going to be mistakes in item writing. Therefore, it is beneficial for every researcher who intends to work on scale development to question themselves in terms of this. On the other hand, that the mistakes are so common in the literature and have become so chronic brings up the need for a control mechanism in the development and the use of psychological measurement tools.

In this framework, it is recommended to establish a national Test Research Implementation Institution that supervises the development, adaptation, revision and use of psychological measurement tools. In addition, researchers who are going to use a psychological measurement tool in their studies are also advised to be careful and careful in choosing the tools they will use.

References

- Acar Güvendir, M. ve Özer Özkan, Y. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi. *Elektronik Sosyal Bilimler Dergisi*, 52, 23-33.
- American Psychological Association, American Educational Research Association & National Council on Measurement in Education. (2014). *Standards For Educational And Psychological Testing*. Washington: American National Research Education.
- Anastasi, A. (1988). *Psychological Testing*. (6thEd.). New York: MacMillan Publishing Co. Inc.
- Bandura, A. (2006) *Guide for constructing self-efficacy scales*. Urdan, T. & Pajares, F. (ed.). Self-efficacy beliefs of adolescents (PB) (Adolescence and education). Information Age Publishing.
- Bandura, A. (1994). *Self-efficacy*. In v. S. Ramachandran (ed.), encyclopedia of human behavior, 4, 71-81. New York: Academic Press.
- Bandura, A. (1997). *Self-efficacy: The Exercise Of Control*. New York: Freeman.
- Conway, Malcolm J. (2006). *How To Collect Data: Measurement And Evaluation*. NY: American Society for Training and Development.
- Cronbach, L. J. (1990). *Essentials Of Psychological Testing*. (5thEd.). New York: Harper Collins Publishers.
- Çüm, S., & Koç, N. (2013). Türkiye'de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi, *Eğitim Bilimleri ve Uygulama*, 12(24), 115-135.
- Dawis, R. V. (1987). Scale construction. *Journal of Counseling Psychology*, 34(4), 481-489.
- Delice, A. ve Ergene, Ö. (2015). Ölçek geliştirme ve uyarlama çalışmalarının incelenmesi: Matematik eğitimi makaleleri örneği. *Karaelmas Journal of Educational Sciences*, 3, 60-75.
- Dirlik, E. M. (2014). Ölçek geliştirme konulu doktora tezlerinin test ve ölçek geliştirme standartlarına uygunluğunun incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(2).
- Doğan, E. M. (2009). *Türkiye'deki psikolojik çalışmalarda kullanılan testlerin psikometrik özelliklerinin incelenmesi: Kültürel açıdan test uyarlama çalışmaları*. Yayımlanmamış yüksek lisans tezi, Muğla Üniversitesi Sosyal Bilimler Fakültesi, Muğla.
- Erkuş, A. (2012). *Psikolojide Ölçme Ve Ölçek Geliştirme-1: Temel Kavramlar Ve İşlemler*. Ankara: Pegem Akademi.
- Erkuş, A. (2016). Ölçek geliştirme ve uyarlama çalışmalarındaki sorunlar ile yazım ve değerlendirilmesi. *Pegem Atf İndeksi*, 0, 1221-1234. doi:10.14527/9786053183563.075.
- Gül Ş. ve Sözbilir, M. (2015). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi. *Eğitim ve Bilim*, 40(178), 85-102. doi: 10.15390/EB.2015.4070
- Hinkin, T. R. (1995). A review of scale development practices in the study in organizations. *Journal of Management*, 21(5), 967-988.
- Slavec, A. ve Drnovsek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review*, 14(1), 39-62.
- Kline, P. (1994). *An Easy Guide To Factor Analysis*. New York: Routledge.
- MEB, (1971). *Sekizinci Milli Eğitim Şurası- Esaslar, Raporlar, Kararlar*, Ankara: Milli Eğitim Bakanlığı Yayınları.
- Özge, S. (1981). *Psychological Tests Used in Turkey: A Preliminary Survey*. Yayımlanmamış yüksek lisans tezi. İstanbul: Bogaziçi Üniversitesi.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Education Research*. 66(4), 543-578.
- Tavşancıl, E., Güler, G. ve Ayan, C. (2014). 2002-2012 yılları arasında Türkiye'de geliştirilen bazı tutum ölçeği geliştirme çalışmalarının ölçek geliştirme süreci açısından incelenmesi. IV. Ulusal

- Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı) 9-13 Haziran*, Hacettepe Üniversitesi, Ankara.
- Tezbaşaran, A. (2008). *Likert Tipi Ölçek Hazırlama Kılavuzu*. Ankara: Türk Psikologlar Derneği Yayınları.
- Tosun, C. ve Taşkesenligil, Y. (2015). The instruments used in science education in Turkey: a descriptive content analysis. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 15(2), 364- 383.
- Zimmerman, B. J. (1995). Self-efficacy and educational development. In: Bandura, A. (ed.), *Self-Efficacy in Changing Societies*. N.Y.: Cambridge University Press.
- Zusho, A.,& Pintrich, P.R. (2003). Skill and will: The role of motivation and cognition in the learning of collage chemistry. *International Journal of Science Education*, 25(9), 1081- 1094.
- Worthington, R. L. ve Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806-838.

Genişletilmiş Özet

Giriş

Duyuşsal özellikler doğası gereği doğrudan gözlenemez ve ölçülemezler. Bu nedenle duyuşsal özelliklerin ölçülmesinde dolaylı ölçme yöntemleri kullanılmaktadır. En sık kullanılan dolaylı ölçme yöntemi ise psikolojik test uygulamalarıdır (Anastasi, 1988). Bu uygulamalar bireyin testteki/ölçekteki maddelere gösterdiği tepkilerin belli bir yonteme göre puanlanması aracılığıyla yapılmaktadır. Sıklıkla tercih edilen psikolojik testler, gözlem ve görüşme gibi diğer bilgi toplama yöntemlerine göre daha kolay uygulanabilmektedir. Psikolojik testlerin objektif olarak puanlanabilmesi, tek bir seferde büyük kitlelerden veri toplanabilmesi, geçerli ve güvenilir gözlemler yapmaya olanak sağlaması en büyük avantajlarından (Conway, 2006; Cronbach, 1990). Alan yazın incelendiğinde araştırmacıların araştırmaları kapsamında ya var olan ölçekleri kullandıkları ya da geliştirdikleri görülmektedir. Araştırmacıları yeni bir ölçek geliştirmeye iten iki durumdan söz edilebilir. Bunlardan biri, araştırmanın amacına hizmet eden bir aracın geliştirilmemiş olmaması; diğeri ise mevcut ölçeklerin güncelliğini yitirmesi ya da istedik özelliklere sahip olmamasıdır. Ulusal ve uluslararası birçok dergide öz-yeterlik ölçeği geliştirme çalışmalarına rastlamak mümkündür. Öz-yeterlik ölçeği geliştirirken diğeri ölçekler gibi genel olarak standart ölçme aracı geliştirme adımlarının dikkate alınması beklenmektedir. Ancak ilgili alanyazın incelendiğinde standart test/ölçme aracı geliştirme aşamalarının takip edilmeden geliştirilmiş birçok ölçeğe rastlanmıştır. Ayrıca geliştirilen bu ölçeklerin çoğunda ölçülen özelliğin (öz-yeterlik) doğasının gerektirdiği ifadeleri içermeyen maddeler bulunduğu gözlenmiştir. Bu şekilde geliştirilmiş ölçeklerin kullanılması ve bu ölçeklerden elde edilen puanlar ile istatistiksel çıkarımlar yapılması bilimin yanlış yönlendirilmesine neden olmaktadır.

Gelecek araştırmalara yanlış dayanak oluşturmaya neden olan ölçeklerden oluşan bu yığına her gün bir yenisi eklenmektedir. Bu sebeple, geliştirilmiş ancak kullanılması mümkün olmayan çok sayıda öz-yeterlik ölçeğinin varlığı, ciddi bir emek ve zaman kaybının göstergesidir. Bu ölçeklerin kullanıldığı çalışmalardan elde edilen sonuçların geçerliği ve güvenilirliği de bilimsel açıdan sorgulanmalıdır. Alanyazında eğitim ve psikoloji alanında geliştirilmiş ölçeklere ait genel incelemeleri içeren çalışmalara rastlanmıştır. Özel olarak öz-yeterlik ölçeklerinin incelendiği bir çalışmaya rastlanmamıştır. Her psikolojik yapı kendine özgüdür ve kendi çerçevesine göre incelenmelidir. Bu nedenle öz-yeterlik ölçeklerinin geliştirilmesi aşamasında var olan sorunları ortaya koymanın, sorunların giderilmesi için önemli bir adım olacağı düşüncesi araştırmanın yapılmasına gereklilik oluşturmuştur. Bu araştırma alanyazında var olan sorunları betimlemesi ve bundan sonraki öz-yeterlik ölçeği geliştirme çalışmaları için de bilgi verici olması açısından önemli görülmektedir. Bu araştırmanın genel amacı, eğitim ve psikolojide kullanılan öz-yeterlik ölçeklerinin geliştirilmesi aşamasında izlenmesi gereken ölçme aracı geliştirme adımlarının ne derece karşılandığını ortaya çıkarmaktır. Bu genel amaç doğrultusunda giriş bilgileri, kuramsal kısım, madde yazımı, geçerlik ve güvenilirlik ve ikinci uygulama olmak beş bölümle ilgili araştırma sorularına yanıt aranmıştır.

Yöntem

Bu araştırma, ölçek geliştirilirken (öz-yeterlik ölçeği) dikkat edilmesi gereken noktalar incelendiğinden nitel bir araştırmadır. Çalışma kapsamında doküman analizi yapılmıştır. Önceden belirlenen kriterlere göre makaleler incelenerek her bir ölçüt için frekans değerleri elde

edilmiştir. Çalışma kapsamında elektronik ortamdaki ulaşılabilen 1997-2018 yılları arasında yerli ve yabancı literatürde yer alan öz-yeterlik ölçeği geliştirme çalışmaları incelemeye alınmıştır. Araştırmaya dahil edilecek çalışmalar seçilirken ölçülmesi hedeflenen öz-yeterlik kavramının eğitim ve psikoloji ile ilişkili olmasına, ve ölçeğin tüm maddelerinin ulaşılabılır olmasına dikkat edilmiş, bu kriterlere uymayan çalışmalar araştırmaya dahil edilmemiştir

Bulgular ve Tartışma

Araştırmanın amaç doğrultusunda giriş bilgileri, kuramsal kısım, madde yazımı, geçerlik ve güvenilirlik ve ikinci uygulama olmak beş bölümle ilgili 1997-2018 yılları arasında geliştirilmiş öz-yeterlik ölçekleri incelenmiştir. Öz-yeterlik ölçekleri incelendiğinde sıklıkla görülen eksikliklerden biri psikolojik yapının işevuruk tanımının yapılmamasıdır. Ayrıca ölçeklerin neredeyse hepsinde dayandırılan ölçme kuramının eksikliği görülmüştür. Madde yazım süreci ile ilgili verilen bilgilere göre araştırmacıların sıklıkla alan yazında bulunan diğer ölçek maddelerinden faydalandığı ikinci olarak da kompozisyon yazdırma ve oradan madde havuzunu oluşturma işlemine başvurdukları belirlenmiştir. Hazırlanan maddelere ilişkin uzman görüşü alma noktasında ise genel eğilimin sadece konu alanı uzmanına danışmak olduğu görülmüştür. Ölçme değerlendirme ve dil uzmanından görüş alan çalışmaların ise oldukça az olduğu belirlenmiştir. Ölçme değerlendirme ve dil uzmanından görüş alan çalışmaların ise oldukça az olduğu belirlenmiştir. Bu bulgular Çüm ve Koç (2013) ve Dirlik (2014)'ün çalışmalarını destekler niteliktedir. Araştırma kapsamında yapılan incelemeler sırasında ölçek geçerliği çalışmalarında sıklıkla karşılaşılan hatanın aynı veri seti üzerinden AFA ve DFA yapılması olduğu görülmüştür. Ayrıca ölçekten madde atıldıktan sonra ikinci AFA'nın da aynı veri seti üzerinden yapıldığı görülmüştür. Madde atıldıktan sonra elde kalan veri seti hala atılan maddenin etkisinin olduğu veri setidir. Bu nedenle madde atıldıktan sonra tekrar veri toplanmalı ve yeni veri seti üzerinden AFA'nın yapılması gerekmektedir. Bu hatalar da araştırmacıların faktör analizi hakkında yeterli bilgiye sahip olmadıklarını göstermektedir. Faktör analizi ile ilgili elde edilen bulgular Kaya Uyanık ve diğ. (2017) çalışmasının bulgusu ile tutarlılık göstermektedir. Yapılan hatalardan en göze çarpanlardan bir diğerinin ise ölçek-anket karmaşasının yaşanmış olmasıdır. Kavram karmaşasının olması ölçeği geliştirenlerin henüz ölçek ve anket ayırımı bilmedikleri anlamına gelmektedir. Ayrıca tespit edilen başka bir hatalı örnekte ise ölçek Türkçe geliştirilmiş fakat derginin yayın dilinden dolayı makale kapsamında İngilizceye çevrilmiştir. Bir dilde geliştirilen ölçeğin maddelerinin başka bir dile çevrilmesi ancak ölçek uyarlama basamakları ile mümkündür. Ölçeğin bu şekilde İngilizceye çevrilip yayınlanması yanlış anlaşılma ve kullanımlara neden olabilecektir. Ölçek geliştirme uzmanlık gerektiren bir alandır. Bundan dolayı ölçek geliştirecek araştırmacıların süreçte kesinlikle ölçme değerlendirme uzmanlarından destek alması gerekir. Bu bilincin yerleşip, sürece dahil edilmeleri sağlandığı takdirde ölçek geliştirme sürecinde yapılan hataların da azalacağı düşünülmektedir. Yapılan incelemeler sonucunda öz-yeterlik ölçeği geliştirme çalışmalarında ciddi sıkıntılar tespit edilmiştir. Bu sıkıntı ölçek geliştirme basamaklarının yeterince iyi anlaşılmadığının bir göstergesi olarak alınabilir. Ölçek geliştirme çalışmaları son yıllarda sayıca artmasına rağmen, niteliğinin hala düşük olması oldukça düşündürücüdür. Ölçek geliştirme basamakları doğru takip edilse dahi incelenen birçok öz-yeterlik ölçeğinde yanlış ifade edilmiş maddelere rastlanmıştır. Ölçek geliştirme bütünsel bir süreçtir. Bu süreç hem geliştirme basamaklarına hakim olmayı hem de psikolojik yapının kuramsal bilgisine hakim olmayı gerektirir. Psikolojik yapının kuramsal bilgisine hakim olunmadığı sürece o yapının

göstergelelerinin tanımlanması doğru yapılamayacak ve sonuç olarak madde yazımında yanlışlar ortaya çıkacaktır. Dolayısı ile ölçek geliştirme çalışması içine girmeye niyetlenen her araştırmacının kendini bu anlamlarda sorgulamasında fayda görülmektedir. Diğer yandan alanyazında hatalı kullanımların bu kadar yaygın olması ve kronik hale gelmesi durumu, psikolojik ölçme araçlarının geliştirilmesi ve kullanılmasında bir denetleme mekanizmasına ihtiyaç olduğu fikrini akla getirmektedir. Bu kapsamda, psikolojik ölçme araçlarının geliştirme, uyarlanma, revizyon ve kullanımlarını denetleyen ulusal bir Test Araştırma Uygulama kurumunun kurulması önerilmektedir. Ayrıca çalışmalarında bir psikolojik ölçme aracı kullanacak araştırmacıların da kullanacakları araçları seçerken dikkatli olmaları ve titiz davranmaları önerilmektedir.