Examination of Motivation Scales: Is the Purpose Academic Promotion or the Need to Measure Psychological Constructs?

Asiye Şengül Avşar, Assoc. Prof. Dr., Recep Tayyip Erdoğan University, asiye.sengul@erdogan.edu.tr
0000-0001-5522-2514

Fulya Barış Pekmezci, Assist. Prof. Dr., Bozok University, fulyabaris@gmail.com
0000-0001-6989-512X

Abstract

The importance of measurement tools in education and psychology is indisputable. It is necessary to measure the affective characteristics of individuals in a valid and reliable manner. Numerous measurement tools have been developed to measure many psychological variables concerning humans. The aim of this research is to examine scale development studies in a technical sense and to determine the usage frequency of the reviewed studies. In line with the determined purpose, 43 motivation scales developed to measure motivation in various fields from the Turkey Index of Measurement Tools database were examined. In this study, which was designed as document analysis, one of the qualitative research designs, the data were examined through the forms developed by the researchers. According to the results, it was determined that there were serious technical errors in the development of measurement tools. It was determined that these errors are mainly in the concepts of test tryout and pilot application, exploratory factor analysis, and confirmatory factor analysis, which seriously affect the psychometric properties of the scales. In addition, when the usage frequencies of the scales were examined, more than half of the scales remained as just development work. Only 19 of the tools examined for motivation were used in other research studies. It is recommended that researchers work in accordance with the standard test development steps in developing scales. Before starting to develop a scale, it is recommended that researchers determine if the scale is really needed with a rigorous literature review.

INTRODUCTION

Psychological tests are most commonly used for measuring affective characteristics, which, due to their nature, are not directly observed and measured (Anastasi, 1988; Baykul, 2000; Özgüven, 2011). Because of their easy implementation, psychological tests are preferred over other data collection methods such as observations and interviews. The most important advantage of psychological tests is that they can be scored objectively with valid and reliable observation and can be collected from a large group at once (Conway, 2006; Cronbach, 1990). One of the most used psychological tests are scales, which are a collection of items and are intended to reveal levels of theoretical variables that are not directly observable (DeVellis, 2016). Within this scope, scales are used for collecting data from a target sample, topic, or content and are able to show the measurement results via mathematical equations.

Researchers develop a scale when they want to measure phenomena that are believed to exist but cannot be assessed directly. Most theories suggest that phenomena exist and influence behavior but that they are intangible. Sometimes, it may be unfeasible to infer the existence of phenomena from behavioral consequences. Therefore, if we are not sure how to interpret samples of behavior, are unwilling to assume behavior as isomorphic, or do not have access to behavioral information, it will be more accurate to assess the construct by using a carefully constructed and validated scale (DeVellis, 2016).

Generally, three motives drive researchers to develop a new scale. One of these is the lack of scales to measure the intended construct; another one is that the existing scales may be out of date, and the last one is that the existing scales do not have sufficient psychometric properties or have many items and are not practical to use. Scale development needs expertise. Nowadays, most researchers develop scales whether or not they have the expertise, and unfortunately, not all scales are developed carefully and decreasing validation. According to DeVellis (2016), for most of them, assembly is a more appropriate term than development. Researchers often gather items and assume that these constitute a proper scale. In this way, researchers may not have an idea whether the items share a common cause that enables the construction of a scale or a common consequence that enables the calculation of an index (DeVellis, 2016).

Scale development requires following standard test development procedures and expertise in the measured construct. Developing a scale without the knowledge of the construct measured or psychometric theory underlying the scale scores is not correct (Erkuş, 2012). According to Crocker and Algina (1986), any doubts regarding the measures of the variables will result in wrong interpretations of analysis concerning those variables, and all conclusions based on them will be false. As a result, measurements will not exhibit the valid values of the related structure, and this will cause measurement error. To obtain a reliable and valid measurement from psychological tests, the scores obtained from the tests should be as free from errors as possible. Providing fewer error-free results is only possible by following standard test development procedures. Developing a test according to test development standards will ensure better quality measurement tools. In this regard, the main purpose of the standards is to enable developed and evaluated scales for the intended implications and to be a guide for interpretations that are made upon the test scores (American Educational Research Association, 2014). According to Crocker and Algina (1986) and DeVellis (2016), the standard test development stages should be as follows:

- Determination of the measured structure
- Determination of the purpose of the scale
- Determination of the theoretical concept and the operational definitions
- Creating an item pool/item writing process
- Determination of how to scale scores (Thurstone, Likert, Gutmann, etc.)
- Obtaining the opinion of experts for the first item pool

- Deciding on the remaining items after the expert opinions
- Implementation of a trial scale on the study group
- Measuring and evaluating the items (item analysis, validity, reliability analysis, factor analysis etc.)
- Second implementation (confirmatory factor analysis)
- Finalization of the scale

Following the steps specified in scale development studies will mean that the psychometric properties of the scales to be developed are satisfactory. This will mean an accurate evaluation of the individuals according to the scores obtained from the measurement tools. In education and psychology, these scores are obtained through the indirect measurement of latent traits. In other words, the responses of individuals to these item groups that reunited theoretically and statistically are considered as proof of the existence of the measured attribute in individuals.

There are lots of attribution which are directly or indirectly observed in education and psychology. One of these attribution is motivation. Motivation is one of the most important and most studied concepts for both psychology and education by reason of being a crucial topic in understanding human behavior. The concept of psychological need has been the most important and core topic of humanities for many years. Gagné et al. (2015), in their research, showed that no matter which country they come from, people's common psychological needs are healthy development, engagement, motivation, and well-being. As it is seen, motivation is an important psychological need of people. Most researchers have tried to propose a theory of motivation. As a consequence of these theories, most researchers have claimed that human behavior has three dimensions: choosing an action, persistence in that choice, and effort to sustain that action (Dörnyei, 2000).

In terms of scope and kind of variable, the notion of motivation varies greatly in the psychology literature. In general, motivation can be explained as a function of different processes that affect and direct one's behavior in order to reach the determined goal (Baron, 1983). Therefore, everyone, even the very little amount, must have motivation. If a person has not motivation, it causes a person not to behave or act, which is named as amotivation (Ryan & Deci, 2000). Besides psychologists, educators have also accepted the significant influence of motivation on human behaviors (Ryan & Deci, 2000). Without motivation, a student is highly unlikely to persist in upcoming challenges. The former U.S. Secretary of Education, Bell, stressed the importance of motivation in education with these sentences: "*There are three things to remember about education. The first one is motivation. The second one is motivation. The third one is motivation.*" (Raffini, 1993; as cited in Lumsden, 1999). From these explanations, it would not be wrong to deduce that motivation is important for success in education. In this context, it is expected that measurement tools will be needed to measure motivation. It is possible to encounter many motivation scales which are not developed according to standard test development procedures in both national and international journals. Using scales that are not developed according to these stages can cause misleading conclusions that mislead science.

When the literature is examined, it was seen that scales were mostly developed in the fields of education and psychology. Also, reliability and validity studies of those scales were generally done in the development phase. However, it was seen that most scales were not being monitored to see if it's outdated or not. In this context, there were various studies conducted in terms of psychometric qualities of scales in national and international literature (Acar Güvendir & Özer Özkan, 2015; Barış Pekmezci & Ayan, 2020; Çüm & Koç, 2013; Delice & Ergene, 2015; Doğan 2009; Erkuş, 2007; Gül & Sözbilir, 2015; Güler & Ayan, 2020; Hinkin, 1995; Slavec & Drnovsek, 2012; Tavşancıl et al., 2014; Tosun & Taşkesenligil, 2015; Worthington & Whittaker, 2006). In these studies, it was stated that most of the new scales were generated regardless of whether they were needed or not. However, there is no study existed about the usage frequency of developed scales which is one of the concerns of this research.

In summary, every day, more and more scales are being developed in that way, creating a false basis for future research studies. Failure to follow the scale development steps or to sufficiently investigate the theory regarding the construct to be measured is effective in the creation of this false basis. Therefore, the accumulation of scales, which are non-qualified and non-scientific, increasing every day. When the literature is examined, a large number of scales developed concerning to measure many psychological constructs can be seen. It is an important issue whether the developed scales are used or not by other researchers. A scale developed according to the scale development steps is expected to give valid and reliable results in different implementations.

The aim of this study is to examine the scales for motivation, which is an important concept in education and psychology. According to the literature review, no study has been encountered which is concerned either with the inspection of motivation scales or the usage frequency of these scales. Accordingly, this research is considered important in determining the qualities of motivation scales in the literature and their functionality calculated with their frequency of use in the field.

From this point of view, because of the importance of motivation both in individuals' educational life and in their social life, the aim of the current study is to examine the development phase of motivation scales and their frequency of usage by other researchers. In line with this, the study has two stages. The first stage involves examining motivation scales and the second stage concerns determining their usage frequency with detailed information. In line with the purpose of the research, the following research questions were examined:

1. When scales are technically analyzed:
    1.1. Has the purpose of the scale been determined correctly?
    1.2. Has the scale been developed in accordance with the theoretical structure/foundation?
    1.3. Was the item writing process appropriate?
    1.4. Has a pilot study been made?
    1.5. Has the test tryout been made properly?
    1.6. Has the analysis (validity and reliability studies) been made sufficiently and correctly?
2. When the frequency of use of the scales is examined:
    2.1. What is the usage frequency of the scales?
    2.2. How many times has the scale been used except by the scale developers themselves?

METHOD

RESEARCH DESIGN

This research is qualitative because the examination of scale development articles related to motivation requires the document analysis technique. The document analysis technique involves the analysis of written materials containing information about the topics to be researched (Patton, 1990). The document analysis aims to find and examine the documents on the subject and to achieve a synthesis that will reveal certain situations or opinions (Maxwell, 1996).

DOCUMENTS

The motivation scale development studies in the national literature between the years 2011 and 2021 which were retrieved from the Turkey Index of Measurement Tools (TOAD) website (https://toad.halileksi.net/) were examined. TOAD is a database developed by Ekşi and Demirci in 2016 consisting of data obtained from open access sources. This database includes scales developed in Turkey or adapted to Turkish culture. As of the date of the research, 9243 scales are registered in the database. Within the scope of the research, by determining the keyword "motivation", 43 articles developed for measuring motivation in various fields were accessed. Only scale development studies related to motivation from any discipline were examined. The reviewed studies were given in the Appendix B.

DATA COLLECTION TOOLS AND PROCEDURE

In the scope of this research, articles were examined via prepared forms. To examine the scales in line with the purpose of the research, the "Technical Scale Evaluation Form (TSEF)" and "Form for Usage Frequency of the Scales (UFS)" were developed by the researchers. In the development of the TSEF, the steps to be followed during the scale development stages (Crocker & Algina, 1986; Cohen & Swerdlik, 2009; DeVellis, 2016) and the measurement tool used in the research by Barış Pekmezci and Ayan (2020) were used.

The TSEF and UFS were presented to two experts in measurement and evaluation and one Turkish language expert. According to the opinions and suggestions of the experts, the forms were finalized. Motivation scales were examined under five headings in the TSEF, which determined the purpose of the scale, theoretical background, item writing (with pilot study), test tryout, and analysis (validity and reliability studies). To determine whether the standard test development procedure was followed, the scale examination form was created according to these five headings. The TSEF consists of 26 items with three categories "yes", "no", and "partially". TSEF was given in Appendix A.

In the TSEF, the section on the purpose of the scale is where the need for a new scale is put forward. The theoretical background of the scale is the section where the theoretical structure to be measured, which is an important issue in scale development, is described, and the operational definition of this structure is made. The item writing section is the section where information on the item development processes is included. In this section, it is determined whether statistical evidence, especially regarding expert reviews, has been presented. A pilot study implementation is an important step that should take place in the scale development stages. Information on the pilot study application is inquired about in the item writing section of the TSEF. Unlike the pilot study application, it is determined in the test tryout section whether the scale items were applied to a similar group to the one they will be applied to. Finally, the analysis section is the section where information about the analysis processes made in scale development is included. In particular, information on the exploratory and confirmatory factor analysis processes and reliability estimations of the scale are presented in this section.

The UFS, which is the other form used in the study, consists of the name of the scale, the usage frequency of the scale, user information of the scale, and the titles of the articles using that scale. The main purpose of the development of this form is to exhibit the usage frequency of the developed scales.

The ethical commission clearance for the research was obtained at the meeting of the Ethics Committee of Recep Tayyip Erdoğan University dated 20.04.2021 and numbered 2021/104.

DATA ANALYSIS

A qualitative approach has been adopted in this research. The data were analyzed by descriptive analysis. The descriptive analysis includes the processes of interpreting the data obtained in qualitative research by summarizing them according to pre-defined themes (Patton, 1990).

The selected articles in this research were randomly distributed among two researchers. The consistency between the coding of the articles was determined according to the form developed by the researchers. For this, the researchers interchanged and recoded 10 articles. Inter-coder consistency or inter-coder reliability was determined by Equation 1 (Miles & Huberman, 1994).

$$Consistency = \frac{Number\ of\ agreement}{Number\ of\ agreement + Number\ of\ disagreement}\ \text{(Equation 1)}$$

A value of at least 0.70 from Equation 1 is accepted as an indicator of inter-coder agreement (Tavşancıl & Aslan, 2001). In this study, a total of 20 articles selected randomly were examined according to the TSEF. The consistency between raters was 0.89 for the TSEF. For the part where

consistency between coders was not achieved (11%), a consensus was reached as a result of interviews.

FINDINGS

FINDINGS RELATED TO TECHNICAL EXAMINATION OF THE SCALES

As a first step, the articles analyzed according to the principles of scale development steps. In other words, the developed scales were evaluated technically. Therefore, the findings obtained according to the parts of the TSEF were examined. The findings for the purpose and theoretical foundation parts are given in Table 1.

Table 1. *Findings Related to Purpose and Theoretical Foundation*

	Y	N	P	Total
The reason of why a new scale was needed was explained.	20	10	13	43
The purpose of the scale was explained.	37	6	0	43
Target group of scales was described.	31	10	2	43
The theoretical foundation of the psychological construct that is to be measured was explained.	21	11	11	43
The measured psychological construct (motivation) was expressed operationally.	11	18	14	43
The measurement theory in which the measurement tool was developed was specified.	0	43	0	43

Y: Yes, N: No, P: Partial

Table 1 indicates that the reason why a new scale was needed was explained in 20 articles. In 13 articles, the reason why the scale was needed was partially explained, but in 10 articles, there was not any explanation. In general, the purpose of the scale was explained in the articles, and the target group to which the scale would be applied was clarified. It was determined that in half of the articles, information was given about the theoretical foundation of the psychological construct intended to be measured (motivation in this study), and in very few of them (11) an operational definition was made.

The data obtained in the scale development studies should be collected and analyzed appropriately according to the test theory determined by the researchers. However, no information was given about the test theory with which the measurement tool was developed in any article examined in this study. Table 2 shows the findings for item writing.

Table 2. *Findings Related to Procedure of Item Writing*

	Y	N	P	Total
Information was given about the item writing.	38	1	4	43
The items were examined by experts.	32	3	8	43
Content validation ratio (CVR) and Content validation index (CVI) values were calculated for item validity.	5	38	0	43
Understandability of the items was tested in a small group (pilot study).	22	21	0	43

Y: Yes, N: No, P: Partial

When Table 2 is examined, it is seen that in most of the studies, information on item writing was given and that experts examined the written items. However, in general, the statistical evaluation of expert opinions was not made. In other words, experimental evidence for expert opinions was not provided with CVR or CVI coefficients. In some of the articles, to obtain opinions, experts were chosen incorrectly. For example, in one of the studies, teachers are considered experts, while in another study, details about the experts' detail did not exist. Also, experts were mentioned as educational scientists, and in most of the studies, the opinions of measurement and evaluation experts were not obtained.

An important step in item writing is testing the understandability of the written items in a small group. This application is called a pilot study (Crocker & Algina, 1986). The group chosen for this application should be similar to the target group for which the scale is planned to be implemented. In this implementation, it should be observed whether the group answering the items did not understand or did not hesitate while reading the items (Crocker & Algina, 1986). Group members should be discussed the items. However, it was seen that the items were tested in a small group in half of the examined articles. Also, in the aforementioned articles, detailed information about this process was not given. Findings related to testing tryout implementation are given in Table 3.

Table 3. *Findings Related to Test Tryout Implementation*

	Y	N	P	Total
The tryout group was similar to the target group.	7	0	0	7
The sample size of the tryout group was large enough.	1	6	0	7
The tryout group data were analyzed properly.	6	0	1	7

Y: Yes, N: No, P: Partial

After clarifying the understandability of the items, a test tryout application should be performed. For the test tryout application, participants with similar characteristics to the target group should be reached. It was concluded from the investigations of the articles that the researchers could not seriously distinguish between the test tryout and pilot study implementation. The number of studies that made test tryout applications was limited to seven. The researchers generally tested the understandability of the items in the test tryout implementation part. Yet this process is known as the pilot study. Also, it was determined that in general, the sample size was not sufficient in the studies in which it was stated that a test tryout application was performed. However, it should be noted that the data were analyzed appropriately in these studies. Findings related to the analysis part are given in Table 4.

Table 4. *Findings Related to Analysis Part*

	Y	N	P	Total
Evidence for construct validity was collected.	41	0	2	43
Factors were extracted according to exploratory factor analysis (EFA).	42	1	0	43
*Principal component analysis (PCA) was used in factor extraction.	31	3	0	34
*Principal axis factoring (PAF) was used in factor extraction.	3	31	0	34
**The justification for the use of PCA was explained.	1	30	0	31
**The justification for the use of PAF was explained.	2	1	0	3
*Orthogonal rotations were made.	33	4	0	37
*Oblique rotations were made.	4	33	0	37
Other validation methods (concurrent validity, discriminant validity, etc.) were used.	5	38	0	43
A single reliability value was calculated.	31	12	0	43
More than one reliability estimation method (test-retest, split-half method etc.) was used.	12	31	0	43
Confirmatory factor analysis (CFA) was performed.	28	15	0	43
***CFA was performed on a new sample.	11	17	0	28
***CFA was performed on the same sample.	17	11	0	28

*Some articles do not have information for the relevant item, **For articles reporting factor extraction methods, ***Results are only for CFA performed studies, Y: Yes, N: No, P: Partial

In scale development studies, there is no doubt that evidence regarding the construct validity of the developed scale should be presented. In this context, it was determined that almost all the reviewed articles collected evidence regarding construct validity. EFA should be performed to determine the clusters of items in scales developed for the first time. In the examined studies, except for one study, EFA was performed. The single study without EFA developed a scale directly by CFA. It is not possible to verify the dimensionality of the existing structure without discovering it. This situation is not considered proper for scale development studies. In the EFA, the aim is to find the items that give the most valid results by selecting the best ones among the written items.

In most of the reviewed articles (34), the factor extraction method was explained, except for nine articles. While almost all the researchers (31) extracted the factors with the PCA, it was determined that the factors were extracted with the PAF in three articles. In the reviewed articles, data analysis was generally performed on a single sample/study group. As can be seen in Table 3, the authors performed EFA without a test tryout implementation. Also, whether the researchers gave information about why they chose the PCA and PAF methods was also examined. Necessary explanations were made in only one of the 31 articles which used PCA and in two of the three articles which used PAF.

After EFA, factors may need to be rotated to interpret the item clusters. It was seen that factor rotation was made in 37 of the examined articles, except for one article. However, in five of the examined articles, no information was available on whether factor rotation was performed or not. Generally, orthogonal rotation (33) was used in rotating the factors. In a limited number of scale development studies (4), oblique rotation was used.

In addition to these results, it is seen that most researchers did not know how to select or exclude items from the scale. Mostly, items were excluded from the scale simultaneously. This is the most alarming method of scale construction because items that are excluded may have affected each other and perhaps just one item among them is responsible for lower factor loadings. When researchers exclude more than one item simultaneously, construct validity will be weakened. Among the examined studies, an extreme example of item elimination was a study in which 20 items were excluded from the scale at the same time. Besides, in one study, an item was excluded from the scale because of having the lowest loading on the scale, which was 0.53. This is not sensible and scientific way because 0.53 is an adequate factor loading for an item to be involved in a factor. Moreover, the problems are not just about item selection, but also related to factor rotations. In one study, it was seen that evidence was provided for non-zero factor correlations, but still, varimax rotation (one of the orthogonal rotations) was used. In another study, one of the factors consisted of just two items. In factor analysis, items are operational definitions of factors and represent a sample of variables that measure the same factor. The rule that factors should consist of at least three items is a sufficient and necessary condition for identification (Anderson & Rubin, 1956; Hair et al., 2019; Rindskopf, 1984). Furthermore, in order not to see Heywood cases (negative error variances) in the structural equation model, the three-items-per-factor rule should be followed (Hair et al., 2019).

Evidence regarding the validity of scores obtained from a newly developed measurement tool should be collected by more than one method. In most of the scale development studies examined (38), only factor analysis results were evaluated for construct validity. In reliability estimates, generally (31) Cronbach alpha coefficient obtained from a single application was interpreted. However, in some studies, it was determined that additional reliability evidence was collected by the test-retest method.

An important step in scale development studies is CFA. CFA application should be made in a new group that has the properties aimed to be examined in the scale, except for the group to which the EFA is applied, in order to determine the validity of the measurement tool that is developed (Henson & Roberts, 2006). It was determined that CFA was made in a considerable number of the examined

scales (28). It was seen that 17 of these studies carried out the CFA on the EFA group. In addition, in 11 studies, CFA was carried out on a new group.

FINDINGS FOR THE FREQUENCY OF USE OF THE SCALES

One of the aims of this study is to determine the usage frequency of the motivation scales developed in different fields. This purpose of the research is important in determining the functionality of the developed scales. Table 5 shows the findings regarding the usage frequency of the scales obtained through the UFS.

Table 5. *Findings Related to Usage Frequency of Scales*

Article number	Year	Citation to self	Citation year to self	Number of citations done by others	Citation years done by others
ID-1	2017	-	-	1	2019
ID-2	2019	2	2019	2	2020
ID-3	2016	-	-	2	2018, 2019
ID-4	2017	-	-	2	2019
ID-5	2017	-	-	1	2019
ID-6	2017	-	-	1	2020
ID-7	2019	-	-	1	2020
ID-8	2018	-	-	2	2019, 2020
ID-9	2018	-	-	6	2019, 2020
ID-10	2018	1	2019	1	2020
ID-11	2018	1	2019	2	2020
ID-12	2018	-	-	1	2018
ID-13	2015	-	-	2	2019
ID-14	2014	2	2014, 2017	3	2016, 2018
ID-15	2014	-	-	4	2018, 2020
ID-16	2013	1	2019	9	2017, 2018, 2019, 2020
ID-17	2013	1	2020	-	-
ID-18	2012	-	-	3	2013, 2018, 2019
ID-19	2015	1	2015	1	2020

When Table 5 is examined, it is determined that 19 of the 43 scales were used by other researchers. When the usage frequency of these measurement tools was examined, the developers of the scales used their own scales in nine different studies. In addition, 19 scales were used in 44 different studies, apart from the researchers who developed the scales. The years when scale developers developed and used the scales are often very close to each other. However, when Table 5 is evaluated, in general, only 18 of the 43 examined scales were used by other researchers.

DISCUSSION, CONCLUSION, AND IMPLICATIONS

In this study, the aim was to examine scales developed for motivation, which is an important psychological concept in education and psychology. This examination was carried out both technically and by determining the usage frequency of the scales. When the results obtained from the technical examination of the scales are evaluated, in general, it was determined that the researchers did not have sufficient backgrounds in scale development.

Scale development starts with determining why a new scale is needed. Various reasons can be expressed for this situation. Examples of these situations are that the existing measurement tools are

not up-to-date, the psychometric properties are not good enough, or the number of items is high. It was seen in the articles reviewed in this study that the reason for the need for a new scale was not sufficiently specified.

Perhaps one of the most important steps in developing scales is to well define the psychological structure to be measured and develop items according to the operational definition. An operational definition is very important in determining the indicators of behavior. This step is directly related to the construct validity. The findings obtained within the scope of this research showed that this important step was seriously ignored.

Scale development can be considered as a scaling of the scores obtained from measurement tools. This requires the developers of the scale to know the test theory they work with and the limitations of the theory. The test theory with which scales were developed was not mentioned in any article in this research. Also, in the literature, some articles concerned with examining scale development have reached the same conclusion (Barış Pekmezci & Ayan, 2020; Çüm & Koç, 2013). In the examinations made, it was determined that scales were developed in the context of classical test theory.

The item writing phase is the step in which both logical and experimental evidence should be provided in scale development studies. The items written are indicators of the psychological structure that is intended to be measured. The more these indicators are written by the structure, the higher the construct validity of the measurement tool will be so that the measurements made according to that measurement tool will serve for the correct evaluation decisions. In general, information was given about the item writing, and experts examined the articles. However, it is questionable whether the experts consulted were suitable for the constructed scales.

Some important points about expert reviews should be highlighted. Domain experts, measurement and evaluation experts, and language experts should study together in scale development processes. In the articles reviewed within the scope of this research, the number of scale development studies involving measurement and evaluation experts is quite limited. This finding is supported by the studies by Barış Pekmezci and Ayan (2020) and Çüm and Koç (2013). Examination of the items by experts is an important step that provides logical evidence about the validity of the items. However, it is important to provide experimental evidence with CVR or CVI coefficients. It was determined that experimental/statistical evidence was not presented in almost all of the articles examined.

It was determined that the pilot study and test tryout stages were confused in the examinations made by researchers. The pilot study is a stage in which the clarity of items is determined through a small group. Test tryout includes pre-item analysis and pre-factor analysis studies. This is done before the actual application. Preliminary analyses are performed on the data obtained. It was determined that this stage was largely ignored in the studies examined.

In the examinations made, it was determined that there is a perception that providing evidence for construct validity is equal to doing factor analysis. Moreover, the factor analysis carried out also contains important problems. Almost all of the studies used PCA and the orthogonal rotation technique. It was not generally stated why PCA was used for factor extraction. PCA and PAF are different factor extraction methods from each other. While it is stated that PCA is not a full factor analysis (Field, 2009, p. 638), it is stated that PCA can be used in the preliminary stage of scale development processes (Hair et al., 2019). In addition, PCA and PAF results do not differ in measurement tools which have construct validity in general (Hair et al. 2019, p. 140). When the factor rotation methods were examined, it was determined that there was generally orthogonal rotation. Orthogonal rotation and oblique rotation are different solutions offered to different problems. Generally, rotating the factors gives researchers results that are easy to interpret. Orthogonal rotation

is a rotation technique that does not allow for correlation between factors, while oblique rotation allows for correlation between factors (Field, 2009, p. 644).

The assumption that there is no relationship between factors in social sciences, and the orthogonal rotation made accordingly, is not realistic (Howard, 2016; Tabachnick & Fidell, 2014). Therefore, more attention should be paid to the constructs obtained as a result of orthogonal rotation. In the examinations, only one of the limited researchers who made oblique rotation stated the reason for oblique rotation. These findings revealed that the researchers did not sufficiently internalize the purpose and basic logic of factor analysis and that they did not know the theoretical infrastructure of factor analysis sufficiently.

Another important finding obtained as a result of the investigations is that the researchers accepted that they achieved sufficient results to provide construct validity by obtaining a dominant factor. This is an important mistake. Measured structures do not have to have a dominant factor. Correlated traits (multidimensional item response theory models) are an example of this (Reise, 2012; Toland et al., 2017).

In addition, factor analysis results obtained for data obtained from a single sample or study group are not sufficient for construct validity. With the development of alternative test theories, different evidence for construct validity can be presented. Gudergan et al. (2004) analyzed data using factor analysis, Rasch analysis, and Mokken analysis in their scale development study. They stated that the interpretation of factor analysis results alone would raise a question mark. Also, investigating the validity of similar scales with previously developed scales measuring the same structure or conducting discriminant validity studies is important in terms of providing evidence for the construct validity of scales developed for the first time.

Reliability estimates are generally made with the Cronbach alpha reliability coefficient. Various criticisms are made for this coefficient. In literature, the alpha coefficient has been criticized due to its assumptions. The alpha coefficient reflects unbiased reliability when the items are equivalent (means, standard deviations, and covariances are equal), tau-equivalent (means and covariances are equal, standard deviations are different), or essentially tau-equivalent (means and standard deviations are different, covariances are equal), which means all item covariances are equal (DeVellis, 2016; McDonald, 1999). Therefore, reliability estimates according to alpha reflect a lower bound of reliability. When the item covariances are not equal, a congeneric model, which means that the standard deviations and covariances of items are different, is assumed. McDonald's omega is one of the reliability coefficients which are used in congeneric measurements. Because of the aforementioned reasons, different reliability estimations (e.g., McDonald's omega, Guttman's lambda, composite reliability) should be made.

When a scale development study is designed, one of the most important steps in the process is CFA studies. An accurate CFA is made by testing the developed scale on different groups. CFA performed on the same sample is not technically correct. Henson and Roberts (2006) stated in their research that using the same data set in EFA and CFA is potentially misleading. After item elimination from the scale, neither the second explanatory factor analysis nor the CFA should be made on the same data set. The data set after item elimination still has the effect of the eliminated items. For this reason, new data should be collected after even just one item is eliminated from the scale (Barış Pekmezci & Ayan, 2020). Both EFA and CFA are based on correlation. Performing CFA on the same data may mean that EFA results are verified with a high probability. According to Finch (2020), because the sample data has already resulted in a specific factor structure via EFA, the CFA model is likely to obtain a good fit spuriously. Therefore, the developed scales should be applied in a different group and the evidence regarding the construct validity should be presented with CFA. In the examined studies (Barış Pekmezci & Ayan, 2020; Güler & Ayan, 2020; Kaya Uyanık et al., 2017), it was determined that EFA and CFA were frequently performed on the same group in the scale development studies.

The technical examinations show that researchers have significant shortcomings in some areas in scale development. The availability of ready-made and easy-to-apply statistical software has increased the usage of factor analysis mechanically. However, psychological origin and psychometric subtleties were not taken into account in the context of the scales examined in this study.

The other purpose of this research was to determine the frequency of use of the developed measurement tools. This purpose is important basically for two reasons. The first is to determine the functionality of the developed scales. In other words, is the scale worth improving? Secondly, have these developed scales contributed to the field other than for academic publication and related academic promotions for their developers?

In the examinations, more than half of the scales (25) remained only as a scale development study. In other words, these scales have never been used. Therefore, it is thought that many scales have been developed only for academic promotion. This causes an accumulation of scales in the scientific literature, which is one of the concerns of this research. However, it was determined that 18 scales were used in 43 studies other than by their developers. Based on these values alone, it can be said that there are roughly three studies per scale. The question of whether the scales developed here are really needed comes to mind. Scales are not disposable or should not be developed for single use.

The results obtained from this study showed that the researchers had technical deficiencies in developing scales. The standard steps in scale development studies should be followed appropriately. Care should be taken in steps such as appropriate EFA, factor extraction techniques, and test tryout application. Perhaps the most important issue is to determine whether that scale is really needed before developing a scale. Disposable scale development efforts will cause a waste of time and effort.

Apart from the motivation structure determined in this study, there are many psychological structures aimed at recognizing and understanding human and human behavior. There are many scales developed to measure these structures (see https://toad.halileksi.net/). It is recommended to conduct similar research on these structures. In addition, only scale development studies were examined in this research. A similar study should be made for scale adaptation studies.

Note: *Scale and test were used interchangeable in this article.*

REFERENCES

Acar Güvendir, M., & Özer Özkan, Y. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarlama konulu makalelerin incelenmesi. *Elektronik Sosyal Bilimler Dergisi, 52*, 23-33. https://doi.org/10.17755/esosder.54872

Anastasi, A. (1988). *Psychological testing*. (6th Ed.). New York: MacMillan Publishing Co. Inc.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Eds.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 5, pp. 111–150). California University Press.

American Educational Research Association. (2014). *Standards for educational and psychological testing*. American Educational Research Association American Psychological Association National Council on Measurement in Education.

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

Baron, R. A. (1983). *Behavior in organizations.* New York: Allyn and Bacon, Inc.

Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması.* Ankara: ÖSYM Yayınları.

Cohen, R. J., & Swerdlik, M. (2009). *Psychological assessment: An introduction to tests and measurements*. (7th Ed.). McGraw–Hill Primis.

Conway, M. J. (2006). *How to collect data: Measurement and evaluation*. American Society for Training and Development Press.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. USA: Rinehart and Winston Inc.

Cronbach, L. J. (1990). *Essentials of psychological testing.* (5th Ed.). 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. Retrieved from https://www.idealonline.com.tr/IdealOnline/pdfViewer/index.xhtml?uId=5928&ioM=Paper&preview=true&isViewer=true#pagemode=bookmarks

Delice, A., & Ergene, Ö. (2015). Ölçek geliştirme ve uyarlama çalışmalarının incelenmesi: Matematik eğitimi makaleleri örneği. *Karaelmas Eğitim Bilimleri Dergisi*, *3*(1), 60-75. Retrieved from https://dergipark.org.tr/tr/pub/kebd/issue/67216/1049114

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications.

Doğan, E. M. (2009). *Türkiye'deki psikolojik çalışmalarda kullanılan testlerin psikometrik özelliklerinin incelenmesi: Kültürel açıdan test uyarlama çalışmalar*. [Yayımlanmamış yüksek lisans tezi], Muğla Üniversitesi Sosyal Bilimler Fakültesi, Muğla.

Dörnyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualization of student motivation. *British Journal of Educational Psychology*, *70*(4), 519-538. https://doi.org/10.1348/000709900158281

Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar. *Türk Psikoloji Bülteni, 13*(40), 17-25. https://bursa.psikolog.org.tr/tr/yayinlar/dergiler/1031828/tpb134004.pdf

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme-1: Temel kavramlar ve işlemler.* Ankara: Pegem Akademi.

Field, A. (2009). *Discovering statistics using IBM SPSS statistics.* (4th ed.). Sage Publication: London.

Finch, W. H. (2020). Using fit statistic differences to determine the optimal number of factors to retain in an exploratory factor analysis. *Educational and Psychological Measurement, 80*(2), 217-241. https://doi.org/10.1177/0013164419865769

Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., Van den Broeck, A., Aspeli, A. K., ... & Westbye, C. (2015). The multidimensional work motivation scale: Validation evidence in seven languages and nine countries. *European Journal of Work and Organizational Psychology, 24*(2), 178-196. https://doi.org/10.1080/1359432X.2013.877892

Gudergan, S., Mathies, C., Kyngdon, A., & Kozicki, S. (2004, December). Negotiation style measurement scale development and testing. Paper presented at the Australian and New Zealand Marketing Academy Conference, Wellington. https://opus.lib.uts.edu.au/handle/10453/3133

Gül Ş., & Sözbilir, M. (2015). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi. *Eğitim ve Bilim, 40*(178), 85-102. http://dx.doi.org/10.15390/EB.2015.4070

Güler, G., & Ayan, C. (2020). Review of attitude scales developed in Turkey between 2002-2018 regarding the scale development process. *Journal of Faculty of Educational Sciences, 53*(3), 839-864. https://doi.org/10.30964/auebfd.658488

Hair, J. F., Black, C. W., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis.* (8th Ed.). Annabel Ainscow: United Kingdom.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393-416. https://doi.org/10.1177/0013164405282485

Hinkin, T. R. (1995). A review of scale development practices in the study in organizations. *Journal of Management, 21*(5), 967-988. https://doi.org/10.1177/014920639502100509

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction, 32*(1), 51-62. https://doi.org/10.1080/10447318.2015.1087664

Kaya Uyanık, G., Güler, N., Taşdelen Teker, G., & Demir, S. (2017). Türkiye'de eğitim alanında yayımlanan ölçek geliştirme çalışmalarının uygunluğunun çok yüzeyli Rasch modeli ile incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology, 8*(2), 183-199. https://doi.org/10.21031/epod.291367

Lumsden, L. (1999). *Student motivation: Cultivating a love of learning*. ERIC Clearinghouse on Educational Management, 5207 University of Oregon, Eugene, OR 97403-5207

Maxwell, J. A. (1996). *Qualitative research design: An interpretive approach*. Thousand Oaks, CA: Sage.

McDonald R. P. (1999). *Test theory: A unified treatment. Mahwah,* NJ: LEA Publisher.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. California: SAGE Publications.

Özgüven, İ. E. (2011). *Psikolojik testler.* Ankara: Pegem Yayınları.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. (2th Ed.). Newbury Park, CA: Sage.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. http://dx.doi.org/10.1080/00273171.2012.715555

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika, 49*(1), 37-47. https://link.springer.com/content/pdf/10.1007/BF02294204.pdf

Ryan, R. M., & Deci, E. L. (2000). When rewards compete with nature: The undermining of intrinsic motivation and self-regulation. In C. Sansone and J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search foroptimal motivation and performance* (pp. 13–54). New York: Academic.

Slavec, A., & Drnovsek, M. (2012). A perspective on scale development in entrepreneurship research. *Economic and Business Review, 14*(1), 39-62. http://ojs.ebrjournal.net/ojs/index.php/ebr/article/view/69/pdf

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Pearson: United States of America.

Tavşancıl, E., & Aslan, E. (2001). *Sözel, yazılı ve diğer materyaller için içerik analizi ve uygulama örnekleri*. İstanbul: Epsilon Yayın.

Tavşancıl, E., Güler, G., & Ayan, C. (2014). *2002-2012 yılları arasında Türkiye'de geliştirilen bazı tutum ölçeği geliştirme çalışmalarının ölçek geliştirme sureci acısından incelenmesi*. IV. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı) 9-13 Haziran, Hacettepe Üniversitesi, Ankara.

Toland, M. D., Sulis, I., Giambona, F., Porcu, M., & Campbell, J. M. (2017). Introduction to bifactor polytomous item response theory analysis. *Journal of School Psychology, 60*, 41-63. https://doi.org/10.1016/j.jsp.2016.11.001

Tosun, C., & 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. Retrieved from https://dergipark.org.tr/tr/download/article-file/17460

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806-838. https://doi.org/10.1177/0011000006288127

Appendix A.

Technical Scale Evaluation Form

	Yes	No	Partially
Purpose of Scale			
The reason of why a new scale was needed was explained.			
The purpose of the scale was explained.			
Target group of scales was described.			
Theoretical Foundations			
The theoretical foundation of the psychological construct that is to be measured was explained.			
The measurement theory in which the measurement tool was developed was specified.			
The measured psychological construct (motivation) was expressed operationally.			
Item writing and Pilot Study			
Information was given about the item writing.			
The items were examined by experts.			
Content validation ratio (CVR) and Content validation index (CVI) values were calculated for item validity.			
Understandability of the items was tested in a small group (pilot study).			
Tryout implementation			
The tryout group was similar to the target group.			
The sample size of the tryout group was large enough.			
The tryout group data were analyzed properly.			
Validity and Reliability Studies			
Evidence for construct validity was collected.			
Factors were extracted according to exploratory factor analysis (EFA).			
Principal component analysis (PCA) was used in factor extraction.			
Principal axis factoring (PAF) was used in factor extraction.			
The justification for the use of PCA was explained.			
The justification for the use of PAF was explained..			
Orthogonal rotations were made.			
Oblique rotations were made.			
Other validation methods (concurrent validity, discriminant validity, etc.) were used.			
A single reliability value was calculated.			
More than one reliability estimation method (test-retest, split-half method etc.) was used.			
Confirmatory factor analysis (CFA) was performed.			
CFA was performed on a new sample.			
CFA was performed on the same sample.			

Appendix B.

Article List

Akdemir, E., & Arslan, A. (2013). Development of motivation scale for teachers. *Procedia-Social and Behavioral Sciences*, *106*, 860-864. https://doi.org/10.1016/j.sbspro.2013.12.098

Akyıldız Munusturlar, M., & Munusturlar, S. (2018). Bilgisayar oyunları motivasyon ölçeğinin geliştirilmesi. *Spormetre Beden Eğitimi ve Spor Bilimleri Dergisi*, *16*(3), 81-90. Retrieved from https://dergipark.org.tr/en/pub/spormetre/issue/68402/1076047

Arslan, A., & Taşgın, A. (2019). Ortaokul öğrencilerine yönelik Türkçe dersine yönelik motivasyon ölçeği geliştirme çalışması. *Journal of Computer and Education Research*, *7*(14), 228-249. https://doi.org/10.18009/jcer.565717

Ateş, Ö. T., & İhtiyaroğlu, N. (2019). Yetişkin motivasyon ölçeği: Bir ölçek geliştirme çalışması. *Kastamonu Eğitim Dergisi*, *27*(2), 611-620. https://doi.org/10.24106/kefdergi.2612

Aydın, İ., Demirkasımoğlu, N., Güner Demir, T., & Erdemli, Ö. (2017). Kamu hizmeti motivasyonu ölçeğinin geliştirilmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 50*(2), 105-126. https://doi.org/10.1501/Egifak_0000001399

Aydın, S., Yerdelen, S., Yalmancı, S. G., & Göksu, V. (2014). Biyoloji öğrenmeye yönelik akademik motivasyon ölçeği: Ölçek geliştirme çalışması. *Eğitim ve Bilim*, *39*(176). 425-435. http://dx.doi.org/10.15390/EB.2014.3678

Balantekin, Y., & Oksal, A. (2014). İlkokul 3. ve 4. sınıf öğrencileri için matematik dersi motivasyon ölçeği. *Cumhuriyet Uluslararası Eğitim Dergisi*, *3*(2), 102-113. https://doi.org/10.30703/cije.321344

Demir, G. T., & Hazar, Z. (2018). Dijital oyun oynama motivasyonu ölçeği (DOOMÖ): Geçerlik ve güvenirlik çalışması. *Beden Eğitimi ve Spor Bilimleri Dergisi*, *12*(2), 128-139. Retrieved from https://dergipark.org.tr/tr/pub/bsd/issue/53467/711639

Deniz, H., & Demir, S. (2020). Yazma motivasyonu ölçeği'nin geliştirilmesi: Geçerlik ve güvenirlik çalışması. *Cumhuriyet Uluslararası Eğitim Dergisi*, *9*(2), 593-616. http://dx.doi.org/10.30703/cije.640584

Dündar, S. A. & Çocuk, H. E. (2019). Yabancı dil olarak Türkçenin öğretiminde okuma motivasyonu ölçeğinin geliştirilmesi: Güvenirlik ve geçerlik çalışması. *International Journal of Languages' Education and Teaching, 7*(4), 401-411. http://dx.doi.org/10.29228/ijlet.39665

Ekici, T. (2017). Bireysel ses eğitimi dersine yönelik motivasyon ölçeği geliştirme. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi*, *60*, 337-349. Retrieved from https://dergipark.org.tr/tr/pub/abuhsbd/issue/32973/366504

Ekinci, E., Yenel, F., & Sarol, H. (2012). Doğa yürüyüşlerine katılım motivasyon ölçeği: Geçerlik ve güvenirlik çalışması. *I. Rekreasyon Araştırmaları Kongresi*, 222-228. Retrieved from https://toad.halileksi.net/wp-content/uploads/2022/07/doga-yuruyuslerine-katilim-motivasyon-olcegi-toad.pdf

Eskicioğlu, A. P., & Alpat, Ş. (2017). Ortaöğretim öğrencilerine yönelik kimya dersi motivasyon ölçeğinin geliştirilmesi. *Türkiye Kimya Derneği Dergisi Kısım C: Kimya Egitimi*, *2*(2), 185-212. Retrieved from https://dergipark.org.tr/tr/pub/jotcsc/issue/31858/335678

Faruk, K. (2016). Karaca umre motivasyonları ve etkileri ölçeği. *Journal of Turkish Studies*, *11*(7), 315-330. http://dx.doi.org/10.7827/TurkishStudies.9748

Genç Ersoy, B., & Belet Boyacı, Ş. D. (2018). Sözcük öğrenme motivasyonu ölçeği (SÖMÖ): Geçerlik ve güvenirlik çalışması. *İlköğretim Online*, *17*(1), 255-267. https://doi.org/10.17051/ilkonline.2018.413763

Gömleksiz, M. N., & Kan, A. Ü. (2012). Sosyal bilgiler dersi motivasyon ölçeğinin geçerlik ve güvenirlik çalışması. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, *22*(2). Retrieved from https://dergipark.org.tr/tr/pub/firatsbed/issue/45179/565671

Gözen, E. (2020). Açık alan rekreasyon aktivitelerine katılmayı motive eden faktörlere yönelik ölçek geliştirme çalışması. *Türk Turizm Araştırmaları Dergisi*, *4*(3), 2197-2209. https://doi.org/10.26677/TR1010.2020.475

İleri Aydemir, Z., & Öztürk, E. (2013). Metinlere yönelik okuma motivasyonu ölçeği: Geçerlik ve güvenirlik çalışması. *İlköğretim Online, 12*(1), 2-12. Retrieved from https://dergipark.org.tr/tr/pub/ilkonline/issue/8586/106670

Karabağ Köse, E., Karataş, E., Küçükcene, M., & Taş, A. (2021). Öğretmen mesleki motivasyonu ölçeği geçerlik ve güvenirlik çalışması: Çevrimiçi ve kâğıt kalem uygulamalarının karşılaştırılması. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 51, 479-498. https://doi.org/10.9779/pauefd.707619

Karadağ, M., Zalluhoğlu, A. E., Günal, G. G., Dayıoğlu, O., & Kışla, T. (2019). Derslerde teknoloji kullanımına yönelik motivasyon ölçeğinin geliştirilmesi. *Ege Sosyal Bilimler Dergisi*, *2*(2), 146-157. Retrieved from https://dergipark.org.tr/tr/pub/esbd/issue/49689/619776

Karataş, K., Ardıç, T., & Kaya, İ. (2016). ARCS motivasyon modeli'ne dayalı öğretim ölçeği geçerlik ve güvenirlik çalışması. *Kastamonu Eğitim Dergisi*, *24*(4), 1821-1838. Retrieved from https://dergipark.org.tr/tr/pub/kefdergi/issue/27734/327565

Katrancı, M. (2015). Book reading motivation scale: Reliability and validity study. *Educational Research and Reviews*, *10*(3), 300-307. https://doi.org/10.5897/ERR2014.1998

Kauffman, D. F., Yılmaz Soylu, M., & Duke, B. (2011). Validation of the motivation to teach scale . *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education), 40,* 279-290. Retrieved from http://www.efdergi.hacettepe.edu.tr/shw_artcl-412.html

Kaya, M. F. (2013). Coğrafya öğrenmeye yönelik motivasyon ölçeği geliştirme çalışması. *Eastern Geographical Review*, *18*(30). Retrieved from https://dergipark.org.tr/en/pub/ataunidcd/issue/2458/31322

Kaynak, S., Özhan, M. B., & Kan, A. (2017). Ortaokul öğrencileri için okul motivasyonu ölçeği geliştirme çalışması. *Journal of Turkish Studies, 12*(4), 293-312. http://dx.doi.org/10.7827/TurkishStudies.11336

Kazu, İ. Y., Özdemir, O., & Erten, P. (2016). İşletim sistemleri ve uygulamaları dersine yönelik motivasyon ölçeğinin geliştirilmesi çalışması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *16*(2), 534-546. https://doi.org/10.17240/aibuefd.2016.16.2-5000194941

Kesici, A. (2018). Lise öğrencilerinin matematik motivasyonunun matematik başarısına etkisinin incelenmesi. *Ondokuz Mayıs University Journal of Education Faculty*, *37*(2), 177-194. https://doi.org/110.7822/omuefd.438550

Kurnaz, H. (2019). Okuma iç motivasyonu ölçeği: Geçerlik ve güvenirlik çalışması. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, *20*(1), 234-250. https://doi.org/10.29299/kefad.2018.20.01.008

Mehdiyev, E., Uğurlu, C. T., & Gonca Usta, H. (2017). İngilizce dil öğreniminde motivasyon ölçeği geçerlik ve güvenirlik çalışması. *The Journal of Academic Social Science Studies*, *54*(1), 21-37. Retrieved from https://jasstudies.com/DergiTamDetay.aspx?ID=3632

Ozgul, I., & Yigit, N. (2017). The development of chorus motivation scale (CMS) for prospective music teacher. *Journal of Education and Learning*, *6*(3), 323-339. http://doi.org/10.5539/jel.v6n3p323

Öz, N. D., & Üstün, F. (2019). E-Spor katılım motivasyonu ölçeği'nin (EKMÖ) geçerlik ve güvenirlik çalışması. *Türk Spor Bilimleri Dergisi*, *2*(2), 115-125. Retrieved from https://dergipark.org.tr/tr/pub/tsbd/issue/48256/590133

Özdemir, E., Kural, M., & Kocakülah, M. S. (2018). Ortaöğretim öğrencilerinin fizik dersine ait motivasyon düzeylerini belirlemeye yönelik ölçek geliştirme. *Kastamonu Eğitim Dergisi*, *26*(5), 1497-1507. https://doi.org/10.24106/kefdergi.2123

Öztürk, E., & Uzunkol, E. (2013). İlkokul öğretmeni motivasyon ölçeğinin psikometrik özellikleri [The psychometric properties of the primary school teacher motivation scale]. *Eğitimde Kuram ve Uygulama*, *9*(4), 421-435. Retrieved from https://dergipark.org.tr/tr/pub/eku/issue/5458/74006

Sarıçiftçi, A., & Köse, H. S. (2017). Koro dersi motivasyon ölçeği geliştirilmesi ve öğrencilerin motivasyon düzeylerinin karşılaştırılmalı olarak incelenmesi. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *9*(22), 207-224. https://doi.org/10.20875/makusobed.333528

Schreglmann, S. (2018). Developing academic motivation scale for learning ınformation technology (AMSLIT): A study of validity and reliability. *Journal of Education and Learning*, *7*(4), 145-153. https://doi.org/10.5539/jel.v7n4p145

Sevim, O. (2019). Yabancı dil olarak Türkçe öğrenmeye yönelik motivasyon: Bir ölçek geliştirme çalışması. *Atatürk Üniversitesi Türkiyat Araştırmaları Enstitüsü Dergisi*, *65*, 567-586. https://doi.org/10.14222/Turkiyat4170

Tahiroğlu, M., & Aktepe, V. (2015). 4. ve 5. sınıf sosyal bilgiler dersi motivasyon ölçeği formunun geçerlilik ve güvenilirlik çalışması. *Journal of Turkish Studies*, *10*(3), 907-932. https://dx.doi.org/10.7827/TurkishStudies.7944

Tekkurşun Demir, G., & Cicioğlu, H. İ. (2018). Fiziksel aktiviteye katılım motivasyonu ölçeği (FAKMÖ): Geçerlik ve güvenirlik çalışması. *International Journal of Human Sciences*, *15*(4), 2479-2492. https://doi.org/10.14687/jhs.v15i4.5585

Tekkurşun Demir G., İlhan, E. L., Esentürk, O. K., & Kan, A. (2018). Engelli bireylerde spora katılım motivasyon ölçeği (ESKMÖ): Geçerlik ve güvenirlik çalışması. *Spormetre Beden Eğitimi ve Spor Bilimleri Dergisi*, *16*(1), 95-106. https://doi.org/10.1501/Sporm_0000000347

Yavuz Tabak, B., Sönmez, E., Yenel, K., & Kan, A. (2019). Öğretmenlerin iş motivasyonu kaynakları: Ölçek geliştirme çalışması. *Sakarya University Journal of Education*, *9*(3), 408-433. https://doi.org/10.19126/suje.461444

Yıldız, V. A., & Taşgın, A. (2020). Öğretmen motivasyon ölçeği: Geçerlik ve güvenirlik çalışması. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi*, *9*(4), 1741-1754. Retrieved from https://dergipark.org.tr/tr/pub/teke/issue/59561/857061

Zeynel, E., & Çarıkçı, İ. H. (2017). Akademisyenlerin mesleki motivasyon algı düzeyini ölçmeye yönelik bir mesleki motivasyon ölçeğinin tasarımı. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, *17*(3), 125-148. https://doi.org/10.18037/ausbd.417286