Development of a Short Form: Methods, Examinations and Recommendations

Hakan KOĞAR *

Abstract

The aim of this review is to explain the methods that can be used when developing short form of a measurement tool and to examine some short form development studies in the field of health sciences literature by taking into consideration the criticisms of short form development studies. It is seen that short form development studies are especially concentrated in the fields of health sciences. The main reason for this situation has been shown that clinicians need fast and reliable measurement tools to reduce the pressure on them. The review results of the 12 articles selected for this research show that there are very few studies that follow the guidelines for short form development. Researchers are advised to develop the short form of the scale by taking into account the criteria mentioned in this study. It is recommended to select measurement instruments which are developed in accordance with ethical rules and have sufficient psychometric properties. Clinical researchers should be aware that the perception that measuring instruments containing less items are less valid does not show the truth. The same psychometric standards are sought for each measurement tools.

Key Words: Short form, scale development, classical test theory, item response theory

INTRODUCTION

Attempts to develop a short form of an existing measurement tool started at the beginning of the 20th century when it was questioned if it was essential to use all the items on Doll's (1917) Binet-Simon intelligence test to measure intelligence. Studies on the development of short forms, the number of which increased in the 1950s, initially focused on the measurement tools used for clinical assessments as an outcome of criticisms made against numerous items on intelligence and ability tests (Levy, 1968). Levy (1968), who examined the short form development studies in that period, criticized these kinds of studies in his study by claiming that studies aiming to produce short forms diverted from their real purposes and became a commonplace academic activity.

Why Short Forms?

The primary aim of studies on the development of short forms in the mid-20th century was for effective use of the time available (Levy, 1968). The aim was to establish a balance between economic use of time and energy and accurate test estimations (Doppelt, 1956). Today, however, there are different reasons underlying efforts to develop short forms. Some of these are as follows: finding the use of short forms convenient in studies involving multiple cultures with multiple variables, saving time by measuring fewer behaviors, the possibility of developing a child form, reaching the goals of selection and placement more quickly and developing a short form having the same validity as that of a long form. Studies on the development of a short form are observed to be more common in the field of health sciences. This is primarily attributed to the health specialists' need for a quick and reliable measurement tool to relieve the pressure they are under (Smith, McCarthy & Anderson, 2000).

* Assoc. Prof. Dr., Akdeniz University, Faculty of Education, Antalya-Turkey, e-mail: hkogar@gmail.com, ORCID ID: 0000-0001-5749-9824

_____

Psychometric Theories Used In Developing Short Forms

Various methods were used in developing short forms in the mid-20th century, some of which are selecting the item set yielding the highest correlation with the long form of the measurement tool, forming an item sample based merely on item statistics, and selecting a factor or factors with the highest validity (Levy, 1968). It was revealed that among these methods, it was the selection of an item sample based on classical item statistics that was used most frequently; in addition to these statistics, some other statistics, such as Guttman's scalogram analyses were also found to be utilized. These methods, the use of which are limited today, as well as other methods that started to be used with the advancements in technology as of 1970, are explained below in detail in association with the psychometric theories they are based on.

Classical Test Theory (CTT)

Classical item statistics

The most important of the classical statistics that go way back to the times when intense interest in scale development studies started in the field of social sciences are item difficulty index, item discrimination index and item total correlation coefficient. The item difficulty index refers to the difficulty level of an item with respect to the ability level of the individuals in a group. According to Henning (1987), an item being too easy or too difficult can indicate that the score distribution is skewed, which may show that the item prepared is not compatible with the ability level of the group. The item discrimination index, the purpose of which is to distinguish a high scoring group from the low scoring group in reference to the total score, is an important index value that determines the place of an item in a scale. As for the item total correlation coefficient, it displays the relationship between the trait the item or the content is testing and the trait that the total score of the test is measuring. Each item score should be associated with the total score. Items that show a high level of relationship with the total score are those items that highly account for the variance in the total score, as in the factor load of a factor analysis. In other words, these items have a high level of validity. These statistical techniques are frequently utilized in the 21st century as their calculations are relatively easy. However, particularly item total correlation is known to result in misleading findings as it is based on the Pearson's product-moment correlation coefficient (Raykov & Marcoulides, 2011)

Biggers' (1976) Spearman-Brown prediction method

Biggers (1976), who criticized the use of classical item statistics, stated that the long form is the unity of n number of parallel short forms, and that the short form developed is merely one of these parallel forms; thus, it is not possible to determine which short form is a more appropriate selection. Moreover, generating a short form by eliminating or choosing items is an irreversible experimental method; that is, he stated that it was not possible to initially develop a short form and then add items to try to obtain the long form of the test. For this purpose, Spearman-Brown proposed the prediction method as an alternative to developing a short form. He, first of all, developed the short form of a 40-item dogmatism scale with the aid of classical item analyses. Subsequently, he divided the test into two parts based on odd-and even-numbered items, and calculated the correlation coefficient between the total score of the short form, obtained using the classical item analyses, and the total score of the long form of the scale. It was found that the coefficient between the scores obtained from one half of the scale based on odd-numbered items and the scores from the long form of the scale was .92, while the correlation between the scores of the other part of the scale based on even-numbered items and the scores obtained from the long form of the scale was .93.

_____

_____

Factor analysis

Factor analysis is a multivariate statistical technique by which items are associated with one or more latent items by means of a model constructed based on the relationships among the observed variables. It is the most frequently used statistical technique in studies on scale development and adaptation as well as in short form scale development studies. However, sample studies in which factor analysis is accurately conducted is rarely encountered. According to Goretzko, Pahm and Buhner (2019), in studies where factor analysis is utilized, problems are experienced particularly in identifying the sample size, in choosing the correct rotation method and the correct technique for selecting the factor-revealing technique. Based on the studies they examined, Fabrigar, Wegener, MacCallum and Strahan (1999) made some recommendations for studies in which factor analysis would be used. According to researchers, the number of items that needs to be included in a factor is at least four, and the sample size needs to be at least 400. In cases where multivariate normality is obtained, mostly likelihood estimation, and in other conditions such techniques as data rotation methods or principal axis factoring should be used. Smith, McCarthy and Anderson (2000) stated that factor analysis was frequently used in short form development studies and criticized the formation of the short form by applying a factor analysis to the data set obtained from the long form of a scale. This kind of an approach is based on the assumption that the long and short forms of a scale have the same structure. However, there is no certainty that the long and the short forms of the scale have the same factor structure. As a solution to this problem, they proposed running a separate factor analysis on the items of the short form. If these findings are similar to those obtained from the long form, then this means that the two forms of the scale can be alternatives to each other. On the other hand, significant differences between the factor structures of the short and long forms can indicate that these two forms measure different traits.

Item Response Theory (IRT)

Item response theory (IRT) was developed to overcome the various limitations of CTT and particularly the inadequate approaches in determining psychometric properties of scales. It includes two approaches, namely parametric (Birnbaum, 1968; Rasch, 1960) and non-parametric approaches (Mokken & Lewis, 1982). Researchers should choose one of these fundamental approaches based on the purpose of the research study and on the extent to which the assumptions are met. When there is a symmetrical relationship between a latent trait and responses to the item, and when uni-dimensionality and a large sample size can be ensured, parametric IRT models can be utilized. On the other hand, when there is an asymmetrical distribution and a small sample size, non-parametric IRT models can be used. It is known that parametric and non-parametric IRT models show resistance to conditions where the unidimensionality of IRT models are violated (Embretson & Reise, 2000; Sodano & Tracey, 2011).

Parametric Item Response Theory

Like factor analysis, techniques based on CTT can obtain information based only on relationships among independent items. Moreover, all the statistical findings obtained are dependent on the sample. The greatest advantage of the item response theory (IRT) is that it eliminates the dependence on the sample by claiming invariance of the item parameters. The standard errors in IRT are calculated separately for each level of latent trait. In this way, the group's fixation to one error value is overcome. This topic is important in terms of the decisions made especially in clinical measurements. IRT obtains information from the items that can distinguish groups with high and low ability. Furthermore, as IRT yields item characteristic curves (ICC) at each trait level and for each dimension, the amount of information necessary to obtain the short form of the scale can be estimated. While it is possible to determine the level of ability with a higher level of certainty with items yielding higher amounts of information, determining ability level with items yielding lower amounts of information is possible with lower level of certainty. The items yielding the highest amounts of information can be selected in accordance with the range of the trait being measured. By selecting the better performing items providing adequate information across different levels of the trait, it is possible to develop a short form with high psychometric properties. In addition, rather than obtaining a single coefficient yielded as in reliability

_____

_____

measuring techniques based on CTT, such as Cronbach alpha, test information functions (TIF) in IRT allows the assessment of the certainty for each level of the structure being measured. Thanks to TIF, ability levels that include high amounts of information and thus include low amounts of error can be determined and, in this way, a high level of local reliability can be obtained (Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991). TIF can be developed by means of ICCs. Hence, in short form development studies, the aim should be to reach the same amount of information that the long form possesses by selecting items yielding high amounts of information.

Non-Parametric Item Response Theory and the Mokken Scale Analysis

Non-parametric Item Response Theory (IRT) is an approach, the use of which has become widespread as of the beginning of the 21st century owing to the very low number of assumptions it has. Its interpretation is also easy for researchers. It is commonly used particularly for exploratory purposes. Like in parametric IRT, ICCs are also obtained in non-parametric IRT. ICCs can be obtained in all kinds of distributions — monotonically decreasing, monotonously non-decreasing, symmetrical or asymmetrical distributions (Meijer & Baneke, 2004). Non-parametric IRT models are categorized into two: Mokken scale analyses and non-parametric regression prediction models. The Mokken scale analysis is the extended probabilistic version of the Guttman scale. It has two approaches, namely the Monotone Homogeneity Model (MHM) and the Dual Monotone Model (DMM). MHM defines the relationship between individuals and items that belong to unidimensional item groups and that have an item response function displaying a latent trait and a monotonic relationship. It is the simplified version of DMM with fewer assumptions. The primary aim of these models is to order items and individuals (Koğar, 2015). Parametric and non-parametric IRT follows the algorithm for simultaneous selection (Lei, Dunbar, & Kolen, 2004).

Ant Colony Optimization

Even though it is not a psychometric theory, the Ant Colony Optimization (ACO), one of the most current and effective techniques developed with the aim of developing short forms, is based on the algorithm of ants' search for food (Dorigo & Stützle, 2004). It is believed that this algorithm, which calculates the shortest route between the ant colony and the food source, can be used in short form development studies. It is modeled by utilizing the Structural Equality Model (SEM). The ACO algorithm aims to reveal the model with the highest compatibility by converging towards the appropriate model. It tries to produce the best short form based on the repetition of this process.

Purpose and Importance of the Research

In the present study, some of the methods frequently utilized to develop the short form of a measurement tool are explained. Even though the number of studies based on developing short forms is quite high and has a long history, discussions in this area continue to exist. Criticisms against studies on developing short forms can be examined from two basic aspects. First, these studies prioritize the validity of the measurement tool over any other property. According to psychometric theories, the validity of a measurement tool is obligatory. Such factors that relate to convenience, such as reducing item numbers or using time more effectively, are of secondary importance. Hence, while developing a short form of a scale, the primary aim should be to obtain a short form that is at least as valid as the long form of the scale. However, it is noticed in literature that there are research studies that divert from this aim. The second criticism is that during the development of the short form of a measurement tool, methodology errors are frequently made, the short form is developed carelessly and imprecisely, and the short form is not compared to the long form. This could be attributed to the limited information regarding the methodology of developing short forms in the literature (Smith, McCarthy & Anderson, 2000). The aim of the current study is to explain the methods that can be used to develop a short form of a measurement tool and to examine the methodology that some studies employed to develop short forms in the literature

_____

ISSN: 1309 – 6575 *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

305

of education and health sciences by taking into consideration the criticisms made against studies on short form development.

What Needs to be Taken into Account in Short Form Development Studies and The Examination of Some Studies

In this part of the study, what needs to be taken into account while developing short forms are explained and itemized based on the studies of Levy (1968), Smith, McCarthy and Anderson (2000), and Hagtvet and Sipos (2016). For the present study, 12 short form development studies published in journals indexed in the ERIC and PUBMED databases between the years 2011 and 2019 were selected. In all of these studies, the aim was to develop a new, short form. The present study examined whether or not the short form in each was developed in accordance with the principles stated below. Identification regarding these studies is presented in Table 1.

Table 1. Identification of the Studies Examined

Source	The Short Form of the Scale
Baiocco, Pallini & Santamaria (2014)	Adolescent Friendship Attachment Scale
Woudstra, Meppelink, Maat, Oosterhaven, Fransen & Dima (2019)	Short Assessment of Health Literacy
Lim & Chapman (2013)	Attitudes Toward Mathematics Inventory
Jenkinson, Kelly, Dummett & Morley (2019)	The Oxford Participation and Activities Questionnaire
Rogers, M. E., Creed, P. A., Searle, J., & Hartung, P. J. (2011)	Physician Values in Practice Scale
Ferrario, Panzeri, Anselmi & Vidotto (2019)	Illness Denial Questionnaire
Morin, Valois, Crocker & Robitaille (2019)	Intellectual Disability Questionnaire
Nimon & Zigarmi (2015)	Work Intention Inventory
Milavic, Padulo, Grgantov, Milic, Mannarini, Manzoni, Ardigo & Rossi (2019)	The Psychology Skills Inventory For Sports
Park & Hill (2017)	Occupational Work Ethic Inventory
Siefert, Sexton, Meehan, Nelson, Haggerty, Dauphin & Huprich (2019)	DSM–5 Levels of Personality Functioning Questionnaire
Bohlmeijer, Klooster, Fledderus, Veehof & Baer (2011)	Five Facet Mindfulness Questionnaire

1. Initially, the long form of a measurement tool should be sufficiently reliable and valid:

When a short form is to be developed, the first step to be taken is to evaluate the reliability and validity values of this scale long form. If a measurement tool is not reliable nor valid, then any short form of this tool will most likely have inaccurate validity and reliability values. Two of the 12 studies examined explained the psychometric traits of the long form in detail. Other studies sufficed by merely reporting reliability coefficients or stating that the long form is valid and reliable measurement tool.

2. If a short form does not have the same psychometric traits as those of the long form of a measurement tool, then it is not a single short form, but one of the alternative short forms:

The item set in a short form should be formed by randomly selecting an item set from the long form of the scale that best explains the structure. The next phase is to make the decision as to whether the short form is an "equivalent" or "exchangeable" form. The "equivalent" short form has the same psychometric traits as those of the long form and, therefore, can be used as an alternative to the long form. The "exchangeable" short form, however, does not possess psychometric traits to the same degree as those of the long form. Hence, in another study replicated with a similar method, it is likely to obtain similar forms. "Exchangeable" short forms generally have a lower validity than the long form. In this case, the researcher should reveal and discuss the different forms, the different factor structures, and the different items or item sets that can be alternatives to this form. Otherwise, this form cannot be an alternative to

_____

ISSN: 1309 – 6575   _Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi_
Journal of Measurement and Evaluation in Education and Psychology

306

the long form. Having fewer items in the short form does not mean a lower level of validity is sufficient. This issue is so important that it cannot be disregarded. In two of the studies examined, it was deduced that the form assumed an "equivalent" nature. The short forms developed in these studies were at least as valid and as reliable as the long form. However, in the remaining ten studies, since there was no sufficient information about the reliability and validity of the long form, no interpretation could be made about these studies.

3. A transition should be made from the population behavior (items in the long form) to the the sample behavior (items to be included in the short form) by ensuring that it reflects the nature of the trait which the measurement tool is measuring:

One other factor that needs attention is related to the selection of items for the short form from the item pool in the long form. The selected items that will make up the sample of the behavior should be able to reflect the population behavior in the long form. This topic is as important as psychometric properties and is related to content validity. A well-explained and well-defined content is a topic of priority that is of vital importance for construct validity. In order to maintain the content domain, not only statistical evidence but also expertise in the field is important in the selection of the items to be included in the sample. Only one of the studies examined was observed to have discussed the content of the long form in detail and took into consideration the content as well as the statistical analyses when choosing items for the short form. In all the other studies, only statistical evidence was taken into consideration.

4. The view that "if the long form of the measurement tool is valid, then its short form is also valid" is wrong:

Even if a short form includes the items in the long form as well, this does not ensure that the short form will be reliable and valid. The short form includes fewer items and less content. From this respect, it is psychometrically at a disadvantage. For this reason, the psychometric properties must definitely be statistically proven. In all the studies examined, statistical evidence was sought for the reliability and validity of the short form.

5. In measurement tools with multiple dimensions, the content and psychometric properties should be analyzed for each dimension:

In structures with multiple dimensions, the psychometric properties of the scale should be examined by associating each item of the scale with the relevant dimension. In this case, evidence should be presented to prove that each dimension is reliable and valid. For example, if item selection is to be made based on item-total correlations, the total score should be the factor score, not the overall total of the measurement tool. It should be ensured that there are at least four items in one dimension. If it is essential to omit one dimension completely from the scale, then the relevant theoretical and statistical foundation should be presented in detail. It should be noted that the lower the number of items are, the the lower the content validity will be. One of the 12 studies examined was disregarded because it had a unidimensional structure. 10 of the remaining studies was found to have taken into consideration the multidimensional structure and run the statistical analyses. Even though the present study had a multidimensional structure, it obtained the proofs for the latent trait by means of the total score of the scale.

6. Evidence regarding reliability should be obtained within the scope of various types of reliability:

Construct validity should be the primary concern in determining the validity of a short form. However, in reporting reliability, different kinds of evidence for reliability such as internal consistency reliability, inter-rater agreement in measurements of behavior, and stability reliability need to be obtained. Reliability is a concept related to error and it is not possible to mention only one type of error in a measurement process. Hence, reliability coefficients that take error into consideration from different perspectives should be used. Only one of the 12 studies that were examined obtained internal consistency and stability reliability coefficients. To this end, the Cronbach alpha and the test-retest reliability coefficients were used. It was observed that in one of the research studies the reliability coefficient was not reported. All the remaining studies were found to have reported the internal consistency reliability coefficient. Eight of these studies reported the Cronbach alpha reliability coefficient, one reported the

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

307

Raykov's maximum reliability coefficient and one reported the person reliability and person discrimination coefficients.

7. The psychometric properties of the short form should be examined independent of the long form:

The short form of a measurement tool is a copy of the long form which displays a high degree of association. However, this high degree of association does not prove that the short form is reliable and valid. The concepts of validity and reliability are not transferrable and transitive. For this reason, the psychometric properties of the short form must definitely be examined independent of the long form and evidence should be reported. The proofs obtained from one independent group should be compared with the reliability and validity proofs of the long form. While half of the studies examined were found to have obtained the reliability and validity coefficients independent of the long form, the other half of the studies remained limited to merely reducing the number of items in the long form.

8. In clinical and behavioral measurement tools, the classification accuracies of the short form should also be examined:

The aim of some clinical measurement tools is to make classifications. The aim should be to refrain from negative classification (diagnosing an individual with a syndrome as having no syndrome) and positive classification (diagnosing an individual without a syndrome as having a syndrome). Thus, proofs independent of the long form should be obtained. An accurate classification and diagnosis by the long form does not guarantee that the short form can serve the same purposes as well. Four of the studies examined can be used for clinical purposes. None of these studies reported any proof for accuracy of classification.

9. That the time saved by developing a short form is meaningful and important should be justified:

One of the concrete aims of developing a short form is to save time. However, as previously mentioned, validity and reliability are more important than time. Hence, the researcher should explain how much time was saved and show that the time saved did not impact the the psychometric properties. On average 40 minutes is needed to fill in a long form with 80 items. Assuming that the short form of such a form would include 40 items, it can be said that 20 minutes will be saved. However, it should be noted that a reduction of 40 items will have negative impacts on the reliability and validity. The degree of these effects should be discussed in the study. One of the studies examined the time to be saved by developing a short form and discussed this by taking into consideration the psychometric properties of the measurement tool. The other studies, however, merely stated that time would be used more effectively.

CONCLUSION

While developing a short form of a measurement tool, one of the greatest misconceptions of researchers is the idea that the reliability and validity of the short form and the original measurement tool are the same. This causes some researchers to disregard psychometric properties such as reliability and validity, and prevents some researchers from paying the necessary importance to this issue. In the development of a short form, the observed number of items decreases. Therefore, the content and coverage are narrowed, which makes it difficult for these two test forms to be alternatives to each other.

The 12 research studies selected for the present study were screened in two important indexed databases in the fields of health and social sciences. The results which the examinations yielded show that the number of studies conforming to the rules of developing short forms is limited. This shows that short form development studies, which have been under discussion since mid-20th century, are still subject to discussion. When the study by Levy (1968) is compared to that of Smith, McCarthy and Anderson (2000), it is true that the examined studies performed a more accurate study. However, in the examined studies the following problems were identified: not reporting a detailed account of the reliability and validity information of the long form of the measurement tool, not paying attention to the fact that the short form must be as reliable and valid as the long form, not being aware of the fact that the concept of "exchangeable" short form emerges in cases where the reliability and validity of the short form is not at the same level as those of the long form, limiting reliability to merely reporting the internal

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

308

consistency coefficients, obtaining the psychometric properties of the short form independent of the long form, not providing a detailed explanation of the content of the long form of the measurement tool, and not obtaining proof regarding the fact that the content of the short form can be generalized to the long form as well. In these studies, there are also deficiencies in terms of not explaining how much time is saved, which is one of the primary aims of developing short forms, and how this impacts psychometric properties. Furthermore, it was observed that in clinical measurement tools, classification accuracy was not tested.

When developing short forms, researchers are recommended to use the long form of the measurement as a starting point and take the criteria mentioned in the present study into consideration. On the other hand, studies aiming to adapt short forms of the measurement tool are not recommended owing to some important points such as the psychometric properties of short forms may not be precise. For this reason, instead of conducting a short form adaptation study, initially adapting the long form of a measurement tool to the related culture, and then developing the short form of the adapted measurement tool is recommended to be a more sound approach.

Particularly from the clinical researchers perspective, it is not sufficient to choose a measurement tool whose short form is already developed merely because it was published in a refereed journal and because it will save more time. Measurement tools that were developed in accordance with ethical principles and have sufficient psychometric properties are recommended to be selected. Clinical researchers should note that the perception that measurement tools with fewer items are less valid is not true. The same psychometric standards should be sought in each measurement tool. Moreover, in selection of a short form, it is recommended that one should critically analyze whether or not the steps outlined in the present study were followed during its developmental process; a short form that has the essential properties can be utilized for clinical or other purposes.

REFERENCES

Baiocco, R., Pallini, S., & Santamaria, F. (2014). The development and validation of an Italian short form of the adolescent friendship attachment scale. *Measurement and Evaluation in Counseling and Development, 47*(4), 247-255. https://doi.org/10.1177/0748175614538060

Biggers, J.L. (1976). An a priori approach for developing short-forms of tests and inventories. *The Journal of Experimental Education, 44*(3), 8-10. https://doi.org/10.1080/00220973.1976.11011528

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M Lord, & R. Novick (Eds.), *Statistical Theories of Mental Test Scores* (pp. 397–479). Reading, MA: MIT Press.

Bohlmeijer, E., Ten Klooster, P. M., Fledderus, M., Veehof, M. , & Baer, R. (2011). Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. *Assessment, 18*(3), 308-320. https://doi.org/10.1177/1073191111408231

Doll, E. A. (1917). A brief Binet-Simon scale. *Psychological Clinic, 11*, 197-211.

Doppelt, J. E. (1956). Estimating the full scale score on the Wechsler Adult Intelligence Scale from scores on four subjects. *Journal of Consulting Psychology, 20*(1), 63. https://doi.org/10.1037/h0044293

Dorigo, M., & Stützle, T. (2004). *Ant colony optimization*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/1290.001.0001

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: L. Erlbaum Associates.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Ferrario, S. R., Panzeri, A., Anselmi, P., & Vidotto, G. (2019). Development and psychometric properties of a short form of the Illness Denial Questionnaire. *Psychology Research and Behavior Management, 12*, 727. https://doi.org/10.2147/PRBM.S207622

Jenkinson, C., Kelly, L., Dummett, S., & Morley, D. (2019). The Oxford Participation and Activities Questionnaire (Ox-PAQ): development of a short form and index measure. *Patient Related Outcome Measures, 10*, 227-232. https://doi.org/10.2147/PROM.S210416

Goretzko, D., Pham, T. T. H., & Bühner, M. (2019). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology, (In Press),* 1-12. https://doi.org/10.1007/s12144-019-00300-2

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology   309

_____

Hagtevt, K. A., & Sipos, K. (2016). Creating short forms for construct measures: The role of exchangeable forms. *Pedagogika, 66*(6), 689-713.

Hambleton, R. K., Swaninathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.

Henning, G. (1987). *A guide to language testing- development, evaluation, Research*. London: Newbury House Publisher.

Koğar, H. (2015). Madde tepki kuramına ait parametrelerin ve model uyumlarının karşılaştırılması: Bir Monte Carlo Çalışması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 6*(1), 142-157. https://doi.org/10.21031/epod.02072

Lei, P. W., Dunbar, S. B., & Kolen, M. J. (2004). A comparison of parametric and nonparametric approaches to item analysis for multiple choice tests. *Educational and Psychological Measurement, 64*, 565–587. https://doi.org/10.1177/0013164403261760

Leite, W. L., Huang, I. C., & Marcoulides, G. A. (2008). Item selection for the development of short forms of scales using an ant colony optimization algorithm. *Multivariate Behavioral Research, 43*(3), 411-431. https://doi.org/10.1080/00273170802285743

Levy, P. (1968). Short-form tests: A methodological review. *Psychological Bulletin, 69*(6), 410. https://doi.org/10.1037/h0025736

Lim, S. Y., & Chapman, E. (2013). Development of a short form of the attitudes toward mathematics inventory. *Educational Studies in Mathematics, 82*(1), 145-164. https://doi.org/10.1007/s10649-012-9414-x

Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: A case seng for nonparametric item response theory modeling. *Psychological Methods, 9*, 354–368. https://doi.org/10.1037/1082-989X.9.3.354

Milavic, B., Padulo, J., Grgantov, Z., Milić, M., Mannarini, S., Manzoni, G. M. ..., & Rossi, A. (2019). Development and factorial validity of the Psychological Skills Inventory for Sports, Youth, version–Short Form: Assessment of the psychometric properties. *PloS one, 14*(8), 1-17. https://doi.org/10.1371/journal.pone.0220930

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous responses. *Applied Psychological Measurement, 6*, 417–430. https://doi.org/10.1177/014662168200600404

Morin, D., Valois, P., Crocker, A. G., & Robitaille, C. (2019). Development and psychometric properties of the Attitudes Toward Intellectual Disability Questionnaire–Short Form. *Journal of Intellectual Disability Research, 63*(6), 539-547. https://doi.org/10.1111/jir.12591

Nimon, K., & Zigarmi, D. (2015). Development of the work intention inventory short-form. *New Horizons in Adult Education and Human Resource Development, 27*(1), 15-28. https://doi.org/10.1002/nha3.20090

Park, H., & Hill, R. B. (2018). Development and validation of a short form of the occupational work ethic inventory. *Journal of Career and Technical Education, 32*(1), 9-28. https://doi.org/10.21061/jcte.v32i1.1588

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmark's Paedagogiske Institut.

Rogers, M. E., Creed, P. A., Searle, J., & Hartung, P. J. (2011). The physician values in practice scale-short form: Development and initial validation. *Journal of Career Development, 38*(2), 111-127. https://doi.org/10.1177/0894845310363593

Sodano, S. M., & Tracey, T. J. (2011). A brief Inventory of Interpersonal Problems–Circumplex using nonparametric item response theory: Introducing the IIP–C–IRT. *Journal of Personality Assessment, 93*(1), 62-75. https://doi.org/10.1080/00223891.2010.528482

Raykov, T., & Marcoulides, G. A. (2011). Classical item analysis using latent variable modeling: a note on a direct evaluation procedure, *Structural Equation Modeling, 18*(2), 315-324. https://doi.org/10.1080/10705511.2011.557347

Siefert, C. J., Sexton, J., Meehan, K., Nelson, S., Haggerty, G., Dauphin, B., & Huprich, S. (2019). Development of a short form for the DSM–5 levels of personality functioning questionnaire. *Journal of Personality Assessment, (In Press)*, 1-11. https://doi.org/10.1080/00223891.2019.1594842

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*(1), 102-111. https://doi.org/10.1037/1040-3590.12.1.102

Woudstra, A. J., Meppelink, C. S., Maat, H. P., Oosterha, ven, J., Fransen, M. P., & Dima, A. L. (2019). Validation of the short assessment of health literacy (SAHL-D) and short-form development: Rasch analysis. *BMC Medical Research Methodology, 19*(1), 122-131. https://doi.org/10.1186/s12874-019-0762-4

_____

ISSN: 1309 – 6575   *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

310