



An Investigation of Wording Effect in Multidimensional Scales

Gül Güler^{a1}

^{a,b} *Trakya University, Faculty of Education, 22030, Edirne, Turkey*

Abstract

This research aimed to psychometrically examine articles published in Turkey between 2018 and 2023, where multidimensional scales developed in social sciences group negative items under a separate factor. Accordingly, attempts were made to identify problems stemming from the wording effect while determining the number of factors in scales. In this context, journals within the scope of the ULAKBIM institute were searched, and studies published during the last five years and accessible in full text in electronic format were preferred. When searching for relevant studies, research studies meeting the following criteria were taken into consideration: conducted in social sciences, with open access permission, published in Turkish language, containing only articles, and sharing all the items fully grouped under factors. In this context, 52 articles providing all items and factors were accessed. According to the study results, none of these articles used dimensionality determination methods other than exploratory factor analysis techniques in determining the structure. Furthermore, the criteria considered in factor extraction were limited to the options offered by statistical software packages. The issue of negative items forming a separate factor together may stem from the wording effect. To gather evidence for the construct validity of the scales developed, the scale developers are recommended to utilize several dimensionality determination methods throughout the process.

Keywords: Dimensionality; wording effect; negative item; scale development; scale bias

1. Introduction

In social sciences, the latent traits attributed to individuals are considered as constructs, and since these constructs cannot be directly observed, they can be inferred based on the responses of individuals in a measurement tool. Researchers in this field often engage in construct validity studies to explain these structures. A trait intended to be measured may inherently be associated with multiple latent traits. Considering the tests employed in social sciences, most of them measure multiple latent traits. For instance, a measurement tool developed in mathematics to measure financial literacy skills can also measure reading comprehension skills. Therefore, it is useful to know whether the intended construct is unidimensional or multidimensional. Considering the purpose of test development and administration, the validity of decisions made about individuals based on test scores will also be affected.

One of the most commonly used techniques to explain the underlying latent traits of constructs is factor analysis. Factor analysis is divided into two types: exploratory and confirmatory analysis. Factor analysis is particularly preferred in scale development

¹ Corresponding author: Gül Güler. ORCID ID: <https://orcid.org/0000-0001-8626-4901>
E-mail address: gulyuce2010@gmail.com

processes. In exploratory factor analysis (EFA) processes, deciding on the number of factors is one of the most important steps (Cosemans & et al., 2022; Finch, 2020). The accuracy or inaccuracy of this decision will also affect the subsequent steps (Kılıç, 2022). Extracting far fewer factors than necessary compresses the variables into a smaller factor space, leading to loss of information, neglect of important factors, and increased error loads (Cosemans et al., 2022). Extracting more factors than necessary may result in splitting factors, constraining interpretations, or identifying unimportant factors (Cosemans et al., 2022, Finch, 2020). In the past and still today, traditional factor analytic methods have inevitably been used as the initial step in dimensionality determination processes (Lee et al., 2023). However, some researchers have emphasized the risk of solely relying on the logic of the factor model when assessing the dimensions of a scale.

In social sciences, it is possible to come across numerous research studies employing EFA in construct determination processes (e.g., Finch, 2020; Henson & Roberts, 2006; Montoya & Edwards, 2021; Schmitt & Sass, 2011). Most of these studies employ standard options in statistical software (e.g., Scree plot or Kaiser Criterion) when determining the number of dimensions. Since selecting a criterion that performs well under all conditions is important, researchers also consider other analyses like parallel analysis and MAP when deciding on the number of dimensions. In addition to the methods used for deciding on the number of factors in scale development processes, the responses given to the items of the candidate scale are also important. The underlying factor that influences the responses given to the items in the scale should be the characteristic intended to be measured by the measurement tool. Otherwise, all other cases indicate validity issues. In the process of developing scales as measurement tools, the way items are worded is one of the significant factors that can influence the construct. Items related to the scale could be worded either positively or negatively. Negative items, or reverse-scored items, have a directionality that contradicts the logic of the construct intended to be measured (Weijters & Baumgartner, 2012). Examples of negative items are often encountered in scale development studies in social sciences. Some researchers have emphasized using negative items to reduce acquiescence bias due to unconditional acceptance, automatic adaptation, or laziness, where individuals might agree without considering the actual content of statements (Podsakoff et al., 2003). However, over the years, some studies have pointed out that negative items could lead to inconsistent responses from individuals, discrepancies with responses to positive items, or logical errors that could arise due to the lack of attention to negative items (Colosi, 2005; Salazar, 2015).

Wording effects are elements that stem from the way items are worded, beyond the relationship between the items arising from measuring the same construct in a measurement tool (Gu et al., 2015). Common wording characteristics in self-report surveys, such as negative wording, ambiguous wording, and types of statements, often explain a certain proportion of an item's variance, distorting the factor structure and parameter estimates (Ou, 2022). Wording effects tend to generate spurious method factors for unidimensional scales, leading researchers to reach false conclusions regarding multidimensionality. Wording effects have been found in many well-known unidimensional scales, including the Rosenberg Self-Esteem Scale (Ou, 2022).

Negative items in measurement tools generally tend to be more correlated with each other (Salazar, 2015). They may form a separate dimension, especially in unidimensional scales

(Merritt, 2012; Salazar, 2015). They can also cause problems in validating the construct in multidimensional scales. This situation has led some researchers to recommend using only positive items in scales (DeVellis, 2003). Psychological scales generally utilize both positively and negatively worded items to reduce various forms of response bias. Acquiescent bias refers to the tendency to choose positive responses regardless of the content of the items. However, some scholars argue that the strategy of including a mix of both positive and negative items can distort the scale factor structure (e.g., negative items loading onto one or more separate factors), leading to a new bias (Sauro & Lewis, 2011; Dodeen, 2015). Research has suggested some ways for dealing with the issue of wording effect, also known as the methods effect problem. In their work, Forsterlee & Ho (1999) proposed various suggestions for dealing with response bias in Likert-type scales. These suggestions include using only positively worded items, using an equal proportion of negatively and positively worded items, and including a few negatively worded items in the rating scale but not using their scores in calculation other than using them for controlling potential response bias. As observed in the literature, studies investigating wording effects using scales measuring different constructs indicate that the direction of positively or negatively worded items affects individuals' responses (Kula-Kartal & Mor-Dirlik, 2021; Salazar, 2015; Weems et al., 2003). The dimensionality of measurement tools could be influenced by the wording effect due to the relationships between items in the measurement tools. Therefore, this aspect should be taken into consideration when determining the construct.

As observed in the literature, scale development studies have been increasing exponentially every year. Incorrect or faulty practices found in the literature in determining the dimensionality of scales would not serve as good examples for new scale development processes and therefore some technical issues might be repeated in other studies. This study aims to review the multidimensional scales developed in Turkey in recent years, examine scale development studies where negative items are grouped under a separate factor, and determine the utilized factor analysis techniques and other scale development steps. The review of these studies is deemed important in terms of shedding light on future research studies. Considering the errors in decisions based on the total scores obtained for the sub-factors of developed scales, the presence of numerous scales where negative items are clustered into a single factor alone indicates a significant waste of effort and time. The absence of such a study in the literature and the belief that revealing the existing issues in this field would be a crucial step towards addressing these problems constituted the rationale for conducting this research. Considering the purpose and significance of the study, the research sought answers to following questions in scale development studies conducted in Turkey between 2018 and 2023:

1. What is the distribution of respective scale development studies according to topics?
2. What is the distribution of the total number of items and the number of positive and negative items in respective scale development studies?
3. What is the distribution of the number of factors in respective scale development studies?
4. What is the distribution of the factor analysis techniques in respective scale development studies?

5. What is the distribution of factor determination criteria in respective scale development studies?
6. What is the distribution of naming factors where negative items are grouped in respective scale development studies?

2. Method

2.1. Research Model

A descriptive study was conducted in this research to examine the structure of multidimensional scale development studies where negative items were grouped under a separate factor and the factor analysis methods employed in these studies. Document analysis was employed within the scope of the research. Document analysis is a procedure that involves reviewing and evaluating printed or electronic documents without the researcher's intervention (Bowen, 2009).

2.2. Study Group

In this study, preference was given to studies that could be accessed in full text electronically by scanning journals at ULAKBİM, an institute established by TÜBİTAK for the provision of national information and document access services. A total of 52 scale development studies conducted between 2018 and 2023 in social sciences were accessed, in which negative items formed a separate factor. Figure 1 presents the distribution of these studies across years.

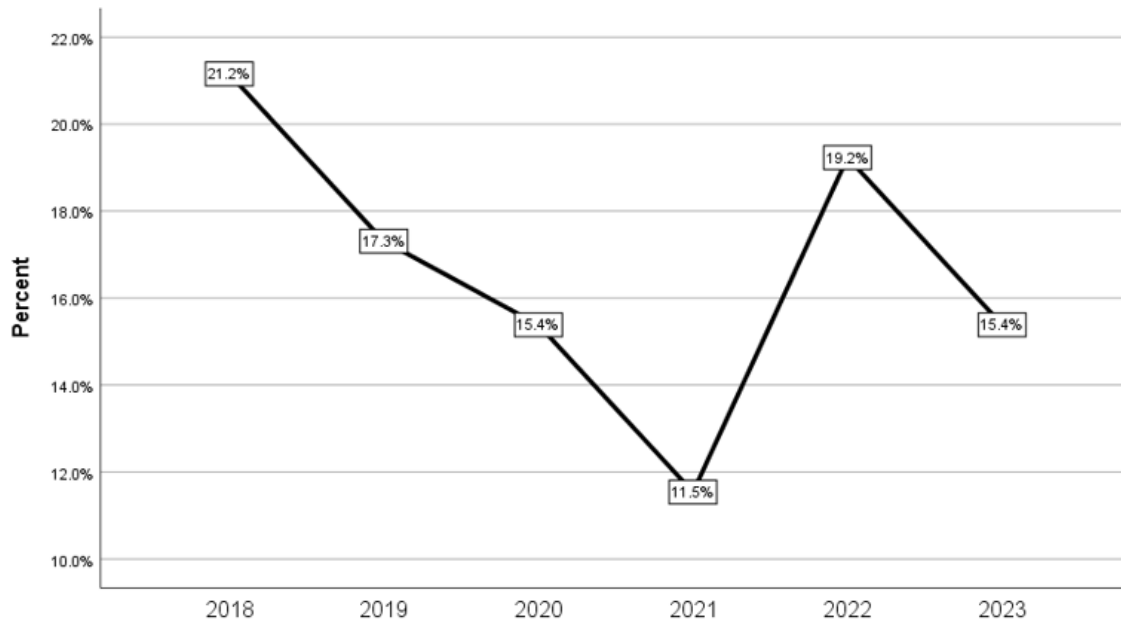


Figure 1. Distribution of examined studies across years

According to Figure 1, 21.2% of scale development studies, in which the negative items were grouped under a separate factor, were conducted in 2018 and 19.2% in 2022. Since 2023 is not over yet, studies published in the first half of the year were taken into consideration.

2.3. Procedure

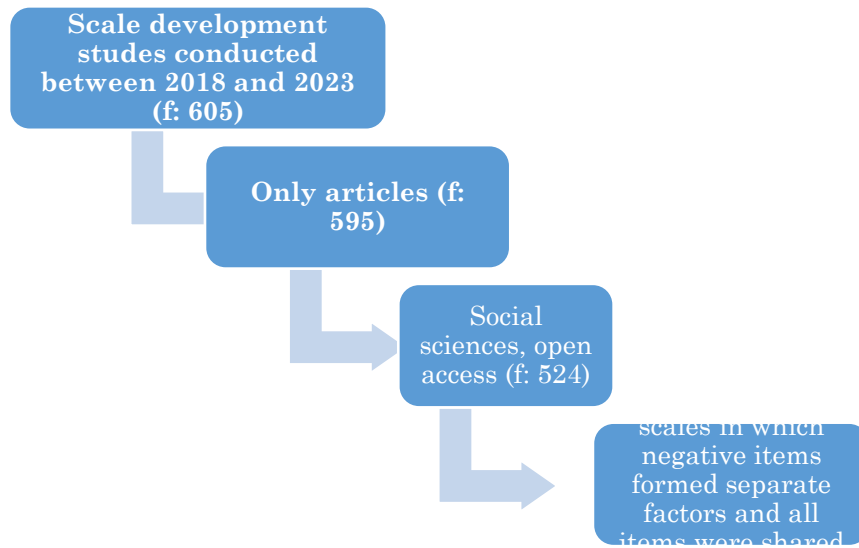


Figure 2. Inclusion criteria for studies included in the study

Articles with open access permission in social sciences, available in TÜBİTAK Ulakbim TR Index, were preferred in the study. The keyword “scale development” was used for searching in the database. The aim was to examine studies conducted in the last 5 years within the scope of the research. However, since we were only in the middle of the year 2023, the year 2018 was also included in the study. Thus, studies conducted between 2018 and 2023 were considered. As such, 605 scale development studies conducted between 2018 and 2023 were accessed. Out of these studies, 595 articles published in peer-reviewed journals were considered, and this number was reduced to 524 open access articles in the field of social sciences. Subsequently, the scale development articles that were taken into consideration were examined. As such, 52 articles were reached where the scales had multidimensional with negative items forming a separate factor, and all the items under the factors were shared in the respective articles. While searching for the relevant studies, only open access articles in social sciences and studies sharing all items grouped under the factors were taken into consideration.

2.4. Data Analysis

The data were analyzed using the document analysis technique. Document analysis has its advantages and certain limitations. The advantages of document analysis include easy online access to numerous documents, the need for selecting data instead of collecting

them, lower costs, and reduced time consumption. The disadvantages of document analysis may include insufficient details in the examined documents, restricted access to documents, or selection bias (Bowen, 2009). In this research, open access articles published in ULAKBIM TR Index between 2018 and 2023 in social sciences were initially examined within the scope of the research questions. Then, the data were entered into the Microsoft Office Excel software considering these criteria, and the frequencies and percentages of the results were reported. In order to ensure coding reliability during the document analysis, 15 out of 52 articles were selected and recoded by the researcher with a one-week interval. To ensure inter-coder reliability, the consensus coefficient was calculated for the 15 studies using the reliability formula suggested by Miles and Huberman (1994) for content analysis studies.

$$\text{Reliability} = \text{number of compromise} / (\text{number of compromises} + \text{number of non-compromises})$$

In this context, the inter-coder reliability was found to be 1.00

3. Results

This section presents the results obtained from document analysis. Table 1 provides the distribution of topics related to the developed scales in the respective studies.

Table 1. Distribution of scale development topics addressed in the articles

Scale development topics	F	%
Related to a psychological construct (tolerance, self-efficacy, coping with bullying, sense of fatherhood, emotion regulation, parenting behaviors, family tranquility, marital role expectations, children's adaptation to divorce, self-belief, group cohesion, self-criticism...)	16	30.8
Related to technology (tablet, computer, online education, EBA, distance learning, digital games, parents' attitudes towards technology, academician attitudes towards emergency distance education due to the COVID-19 pandemic...)	9	17.3
Related to a course (life sciences, social studies, safe laboratory use, dance class, music software...)	8	15.4
Elements related to education (reading, foreign students' learning of Turkish, mathematical reasoning self-efficacy, social studies, critical reading, development-oriented mindset...)	7	13.5
Religious psychology, religious coping, beliefs...	3	5.8
Other (folk dances, rhythm education, workplace satisfaction, environmental problems, tourism perception, cartoons, seasonal agricultural workers, and so on)	9	17.3
Total	52	100

As reported in Table 1, 30.8% of the reviewed multidimensional scales, where negative items formed a separate factor, were related to psychological constructs such as self-efficacy, coping with bullying, sense of fatherhood, self-belief, and self-criticism. In addition, 17.3% of them were related to technology such as tablet, computer, online education, EBA, and digital games. Moreover, 15.4% were related to courses and 13.5% to education. The distribution of topics in Table 1 indicates that most scale development studies in social sciences measure psychological constructs. In addition, there was a considerable number of multidimensional scales related to education with negative items forming a separate factor. Table 2 shows the distribution of the number of factors in multidimensional scales.

Table 2. Distribution of the number of factors in scales

Number of factors	f	%
2	14	26.9
3	21	40.4
4	13	25.0
5+	4	7.7
Total	52	100

As seen in Table 2, 40.4% of the multidimensional scales included three factors, 26.9% included two factors, and 25% included four factors. Considering the wording of the items grouped under factors, the items grouped under one factor in each study consisted of only negative statements, while other factors consisted of positive items related to the respective construct. Table 3 shows the distribution of factor analysis techniques utilized in scale development studies to determine the construct of interest.

Table 3. Distribution of factor analysis techniques utilized in scales

Factor analysis techniques	f	%
Exploratory factor analysis (EFA)	51	98.1
Confirmatory factor analysis (CFA)	48	92.3

As seen in Table 3, considering the distribution of the factor analysis techniques used in scale development studies for determining constructs, 98.1% of the studies employed EFA. However, one study included no information regarding the factor analysis. Also, 92.3% of the studies employed and reported CFA to confirm the structure revealed. At this stage, the criteria considered when determining the number of factors in EFA also important. Therefore, the criteria taken into consideration when deciding whether negative statements grouped under a separate factor due to wording effect truly form a distinct factor in multidimensional scales are presented in Table 4.

Table 4. Distribution of the criteria used when deciding on the number factors in scales

The criteria used when deciding on the number of factors	f	%
K1 rule	46	88.4
Scree plot	29	55.7
Other (Parallel analysis, Comparison data, MAP....)	-	-
Not reported	6	11.5

As seen Table 4, in 88.4% of the scale development studies where negative items formed a separate factor, factors with eigenvalues greater than 1, as proposed by Kaiser (1960) and also known as K1 rule, were considered important factors. Most studies used this rule as a foundation for determining the number of factors. Considering the K1 rule as the only criterion in all EFA studies is not considered sufficient. Indeed, the K1 rule appears to be appropriate only when performing principal component analysis. In 55.7% of these studies, the scree plot was into account as a criterion alongside the K1 rule. Scree plot is a graph that should be used carefully when deciding on the number of factors. In some scale development studies reviewed, it could be observed that instead of the inflexion points in scree plots, only the points were taken into consideration. In this case, relying on points instead of the breaks in the scree plot can mislead researchers when determining the number of factors. It was observed that none of the examined studies used a criterion other than K1 or scree plot. Many recent studies suggest the usefulness of incorporating various criteria such as parallel analysis, MAP, and EGA when determining the number of factors (Cosemans et al., 2022; Golino & Epskamp, 2017; Kılıç & Uysal, 2021). In 11.5% of the examined studies, no information was shared regarding the criteria used for determining the number of factors. Table 5 presents the distribution concerning the naming of factors containing only negative items.

Table 5. Distribution concerning the naming of factors containing only negative items

Naming factors	f	%
Negative thought	4	7.7
Negative attitude	5	9.6
Negative feeling	4	7.7
Negative affective component	2	3.8
Negative factor	5	9.6
Negative dimension	3	5.8
Negative coping	1	1.9
Avoidance	3	5.8
Negative views	2	3.8
Negative beliefs	2	3.8
Other (Blunting, encountered problems, avoidance, destruction effect, affective attitude, destructive self-criticism, fixed mindset, pressure, lack of support, prejudice, difficulty, deep exhaustion, antipathy...).	17	32.7
Not reported	4	7.7
Total	52	100

As seen in Table 5, 9.6% were named as negative attitude, 9.6% as negative factor, and 7.7% as negative thought. The names given to the dimensions where negative items are grouped under a separate factor mostly reflect a negative wording of the characteristic intended to be measured, such as negative dimension, negative factor, negative attitude, negative thought, and so on. In many studies, a dichotomy between positive attitude-negative attitudes, positive thought-negative thought, and similar two-pole differentiations were observed. Table 6 presents the ratio of negative items to all items in the examined scales.

Table 6. Distribution of the total number of items and the number of negative items in scales

Ratios of negative items to total number of items	f	%
Less than half of the total number of items	36	69.3
Half of the total number of items	6	11.5
More than half of the total number of items	10	19.2
Total	52	100

According to Table 6, the number of negative items constituted less than half of the total number of items in 69.3% of the scales, half of the total number of items in 11.5%, and more than half of the total number of items in 19.2%. Approximately every two out of ten scales show that the number of negative items constitute more than half of their total items. Forsterlee & Ho (1999) have suggested various approaches to address the bias caused by working effects in rating scales. They suggested using only positively worded items in rating scales, using both negative and positive items in equal proportions, or using negative items to mitigate potential bias but not including them in the scoring.

4. Discussion

This study examined the scale development studies conducted in social sciences between 2018 and 2023 from a psychometric perspective, particularly focusing on multidimensional scales with negative items grouped under a separate factor. It was observed that multidimensional scales with negative items grouped under a separate factor were most prevalent in 2018 and 2022. However, several multidimensional scales developed especially in the year 2018 did not include the wording of items, resulting in their exclusion from the scope of analysis. Additionally, since the year 2023 has not ended yet, the studies conducted after the start of the research were excluded from the analysis.

Considering the topics of scales reviewed in this research, 30.8% of them were related to psychological constructs such as tolerance, self-efficacy, coping with bullying, emotion regulation, and more. Given the seriousness and importance of the decisions made about individuals based on the results obtained from scales measuring psychological constructs, it is believed that the likelihood of a separate factor emerging due to wording effects in scales could significantly affect the validity of decisions made about individuals after calculating the scores obtained from the scales. Although various studies have investigated the validity problems of using positive and negative items together in psychological rating

scales (Gu et al, 2015; Salazar, 2015; Weems et al, 2003), it would be useful to increase empirical research particularly concerning the measures that should be taken in this regard.

The study found that 26.8% of the examined scale development studies had two factors. Furthermore, factors related to these scales were generally named as positive attitude-negative attitude, positive thought-negative thought, positive factor-negative factor, and the like. Considering the presence of biases stemming from wording effects in well-known unidimensional self-report scales, including the Rosenberg Self-Esteem Scale (Ou, 2022), rerunning factor analysis and gathering new evidence for dimension determination criteria could be useful before calculating scores obtained after the application of these scales in any study.

The scale development studies examined in this research utilized the Kaiser and scree plot criteria for determining dimensionality. It was seen that researchers often did not venture beyond the options offered by statistical software packages. For instance, they did not utilize different suggested dimensionality determination criteria like parallel analysis or MAP methods. For future researchers conducting scale development studies, it would be useful to use more than one criteria when deciding on the number of factors in scales. Recent studies have shown that it is beneficial to explore the use of different criteria such as parallel analysis, MAP, and EGA when deciding on the number of factors (Cosemans et al., 2022; Golino & Epskamp, 2017; Kılıç & Uysal, 2021). For instance, in the laboratory self-efficacy scale development study conducted by Akkuş (2020), the negative items did not form a separate factor since various criteria like parallel analysis were used to determine the number of dimensions.

Considering the ratios of positive and negative items in the scales, they were equal in 11.5% cases, while in approximately two out of every ten examined scales, the number of negative items was more than half of the total number of items. Forsterlee and Ho (1999) have recommended various strategies to eliminate bias caused by wording effects in rating scales. These recommendations include using only positively worded items, using an equal proportion of positive and negative items, and using negative items solely to mitigate potential bias but excluding them from scoring. New studies could explore how the ratios of positive and negative items should be adjusted in scales.

5. Conclusions

This study is limited to studies sharing all scale items. In a different study, the scope of this research could be expanded by obtaining the necessary permissions to include a larger number of studies, particularly focusing on those conducted in 2018 and earlier that did not share the scale items. Lastly, this study includes multidimensional studies with negative items grouped under a separate factor that were conducted in Turkey and indexed in the TR index. Future studies could examine scale development studies found in different databases such as Web of Science (WOS) in terms of the wording effects of items. Additionally, the functionality of negative items in scales could be examined through experimental studies to determine their effectiveness.

References

- Akkuş, A. (2020). Laboratuvar öz yeterlik ölçeği geliştirme çalışması. *Van Yüzyüncü Yıl University Journal of Faculty of Education*, 17 (1), 991-1014. <https://doi.org/10.33711/yyuefd.800917>.
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), <https://doi.org/10.3316/QRJ0902027>
- Colosi, R. (2005). Negatively worded questions cause respondent confusion. Proceedings of the Survey Research Methods Section, American Statistical Association (2005), 2896-2903.
- Cosemans, T., Rosseel, Y., & Gelper, S. (2022). Exploratory Graph Analysis for Factor Retention: Simulation Results for Continuous and Binary Data. *Educational and Psychological Measurement*, 82(5), 880–910. <https://doi.org/10.1177/00131644211059089>.
- DeVellis, R.F. (2003). Scale development: Theory and applications (2nd ed), Sage Publications.
- Dodeen, H. (2015). The effects of positively and negatively worded items on the factor structure of the UCLA Loneliness Scale. *Journal of Psychoeducational Assessment*, 33(3), 259–267. <https://doi.org/10.1177/0734282914548325>
- 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>.
- Forsterlee, R., & Ho, R. (1999). An examination of the short form of the Need for Cognition Scale applied in an Australian sample. *Educational and Psychological Measurement*, 59(3), 471–480. <https://doi.org/10.1177/00131649921969983>.
- Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLOS ONE*, 12(6), 1–26. <https://doi.org/10.1371/journal.pone.0174035>.
- Gu, H., Wen, Z., & Fan, X. (2015). The impact of wording effect on reliability and validity of the Core Self-Evaluation Scale (CSES): A bi-factor perspective. *Personality and Individual Differences*, 83, 142-147.
- Hambleton, R. K., Swaminathan H., & Rogers, H.J. (1991). Fundamentals of item response theory. *California: Sage Publications Inc.*
- Henson, R., & Roberts, J. (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.

- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151. <https://doi.org/10.1177/001316446002000116>.
- Kılıç, A. F., & Uysal, İ. (2021). Faktör çıkarma yöntemlerinin paralel analiz sonuçlarına etkisi. *Trakya Eğitim Dergisi*, 11(2), 926–942. <https://doi.org/10.24315/tred.747075>.
- Kılıç, A. F. (2022). Açımlayıcı faktör analizinde boyut sayısına karar verme: yöntemlere kısa bir bakış, Pamukkale University Sosyal Bilimler Enstitüsü Dergisi, 2022 Sayı 51: Özel sayı 1, Denizli, ss. Ö305-Ö318. <https://doi.org/10.30794/pausbed.1095936>.
- Kula-Kartal, S., & Mor-Dirlik, E. (2021). Examining the dimensionality and monotonicity of an attitude dataset based on the item response theory models. *International Journal of Assessment Tools in Education*, 8(2), 296-309.
- Lee, C., Park Y., & Cho B. (2023). Use of exploratory graph analysis in inspecting the dimensionality of the revised University of California Los Angeles (R-UCLA) Loneliness Scale Among Older Adults. *Res Gerontol Nurs*.16(1):15-20. <https://doi.org/10.3928/19404921-20230104-03>.
- Merritt, S.M. (2012). The two-factor solution to Allen and Meyer's (1990) affective commitment scale: Effects of negatively worded items. *Journal of Business and Psychology*, 27(4), 421-436.
- Miles, M. & Huberman, M.A. (1994). *An expanded sourcebook qualitative data analysis*. London: Sage Publications.
- Montoya, A. K., & Edwards, M. C. (2021). The poor fit of model fit for selecting number of factors in exploratory factor analysis for scale evaluation. *Educational and Psychological Measurement*, 81(3), 413–440. <https://doi.org/10.1177/0013164420942899>.
- Ou X. (2022). Multidimensional Structure or Wording Effect? Reexamination of the Factor Structure of the Chinese General Self-Efficacy Scale, *Journal of Personality Assessment*, 104(1), 64-73. <https://doi.org/10.1080/00223891.2021.1912059>.
- Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903.
- Salazar, M. S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema*, 27(2), 192–199.
- Sauro, J., & Lewis, J. R. (2011). When Designing Usability Questionnaires, Does It Hurt To Be Positive? In Proceedings of CHI 2011 (Pp. 2215–2223). Vancouver, Canada: ACM.

- Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, 71, 95–113. <https://doi.org/10.1177/0013164410387348>.
- Svetina, D. (2011). Assessing dimensionality in complex data structures: A performance comparison of DETECT and NOHARM procedures. Unpublished Ph.D. Thesis, Arizona State University, Arizona.
- Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, 49(5), 737-747.
- Weems, G.H., Onwuegbuzie, A.J., Schreiber, J.B., & Eggers, S.J. (2003). Characteristics of respondents who respond differently to positively and negatively worded items on rating scales. *Assessment and Evaluation in Higher Education*, 28(6), 587-606.
- Zhang, J. (2007). Conditional covariance theory and detect for polytomous items. *Psychometrika*, 72, 69-91.