

A Meta-Analytic Reliability Generalization Study of the Oxford Happiness Scale in Turkish Sample

Vildan ÖZDEMİR *

Yıldız YILDIRIM **

Şeref TAN ***

Abstract

The purpose of this study was to analyze the meta-analytical reliability generalization of short form and long form of the Oxford Happiness Scale (OHS) for Turkish sample. In addition, how different moderator variables affect reliability coefficients was examined. A number of criteria have been set to determine the studies to be included in meta-analysis. Of 95 Cronbach's Alpha coefficients obtained from 92 studies that were selected according to criteria were included in the meta-analysis. In the data analysis, reliability generalization based on meta-analysis was used. The effect of moderator variables on variability in reliability estimations as effect size was examined by Analog ANOVA. As a result of the research, it was found that the mean alpha was .81 for overall studies; .76 for the short form and .87 for the long form of OHS. In addition, it was concluded that number of items had a statistically significant effect on the reliability estimation in terms of heterogeneity of true effect sizes, and sample type had a statistically significant effect on the reliability estimation for OHS (long-form). But sample type had no effect on the reliability estimation for OHS-S (short-form), and field of study had no effect for both short and long form reliability estimates.

Key Words: Reliability generalization, meta-analysis, Oxford happiness scale.

INTRODUCTION

The place and importance of measurement and assessment in education and psychology are indisputable. Accordingly, education and psychology are unthinkable without the field of measurement and evaluation. Two conceptions underlie the field of measurement and evaluation: these are *reliability* and *validity*. The aim of the classical test theory is to present a model to estimate the accuracy of test score measures. And the accuracy is related to the reliability of the test (McDonald, 1999). In short, reliability is the degree of being free from random error of measures. In addition, reliability means consistency of the scores received by the same individuals participating in the same or equivalent tests (Anastasi, 1982). A number of calculations are required to interpret reliability. At this point, the concepts of reliability index and reliability coefficient appear. While the reliability index refers to the relationship between observed scores and true scores, the reliability coefficient refers to the relationship between the scores from the parallel forms. Based on the mathematical relationship between the two concepts, it can be said that the reliability coefficient is the ratio between the true score variance and the observed score variance (Crocker & Algina, 2008). There are formulas suggested by researchers in the calculation of the reliability coefficient. Some coefficients require a single test administration, while others require more than one test administration. One of the most useful characteristics of internal consistency calculations is that it is based on only single test administration (Kline, 2005). Some of the formulas that used in calculating the reliability coefficient in the context of internal consistency are as follows: KR-20, KR-21 (Kuder & Richardson, 1937), Guttman Lambda (λ_3) / or Cronbach's Alpha (α) (Cronbach, 1951; Guttman, 1945), Kristof's coefficient (Kristof, 1963), Stratified alpha (Cronbach, Schönemann, & McKie, 1965), Heise and

* Res Assist, Aksaray University, Faculty of Education, Aksaray-Turkey, vildanbagci@gmail.com, ORCID ID: 0000-0002-9051-8860

** Res Assist, Aydın Adnan Menderes University, Faculty of Education, Aydın-Turkey, yildizyldrm@gmail.com, ORCID ID: 0000-0001-8434-5062

*** Prof. PhD., Gazi University, Gazi Education Faculty, Ankara-Turkey, sereftan4@yahoo.com, ORCID ID: 0000-0002-9892-3369

To cite this article:

Özdemir, V., Yıldırım, Y., & Tan, Ş. (2020). A meta-analytic reliability generalization study of the Oxford Happiness Scale in Turkish sample. *Journal of Measurement and Evaluation in Education and Psychology*, 11(4), 374-404. doi: 10.21031/epod.766266.

Received: 08.07.2020

Accepted: 29.11.2020

Bohrnstedt's Omega (Ω) (Heise & Bohrnstedt, 1970), Armor's theta (θ) (Armor, 1973), Raju's Beta (β) (Raju, 1977), Revelle's Beta (β) (Revelle, 1979), Feldt-Gilmer coefficient (Gilmer & Feldt, 1983), McDonald's Omega (ω) (McDonald, 1985), and Angoff-Feldt coefficient (Feldt & Brennan, 1989).

The reliability and validity of the scores obtained from the measurement tools must be investigated absolutely (Crocker & Algina, 2008). Among the previous coefficients, the Cronbach's Alpha is the most frequently used coefficient in the literature for analyzing and interpreting internal consistency reliability. However, as with all coefficients, Cronbach's Alpha differs from research to research even though the same scale is used because it is a sample dependent coefficient. For example, Cronbach's alpha was .29 in one of the studies in which the Oxford Happiness Scale-Short Form was used and in which the sample was chosen from university students (Taşdibi-Ünlü, 2019), while it was found .97 in another study (İlhan & Güler, 2017). The reliability coefficients vary depending on the variation of the sample characteristics: sample size, administration conditions, time of administration, etc. Such differences in the studies required the generalization of reliability. Reliability generalization (RG) based on meta-analysis was first made by Vacha-Haase (1998). According to Vacha-Haase, the RG analyzes the amount and sources of the variability of the reliability coefficients in different measurements and studies. In other words, the RG study examines whether the reliability coefficient differs between studies. When the literature is reviewed, there are a lot of RG studies based on meta-analysis (e.g. Barnes, Harp, & Jung, 2002; Beretvas, Meyers, & Leite, 2002; Bornmann, Mutz, & Daniel, 2010; Li & Bagger, 2007; Nilsson, Schmidt, & Meek, 2002; Shields & Caruso, 2003; Shields & Caruso, 2004; Vacha-Haase & Thompson, 2011; Vicent, Rubio-Aparicio, Sánchez-Meca, & González, 2019). On the other hand, when the Turkish literature is examined, no RG studies related to a specific scale were found. Some of these studies which have examined different study characteristics that affect the mean reliability estimation in literature can be summarized as follows:

Aguayo, Vargas, Emilia, and Lozano (2011), Capraro and Capraro (2002), and Graham, Liu, & Jeziorski (2006) examined the effect of sample characteristics on reliability estimates, and their results showed that reliability coefficients were dependent on sample characteristics. Also, Caruso (2000) aimed an RG analysis of NEO Personality Scales and examined the effect of sample characteristics. He founded that there was a significant difference between the reliability coefficients for sample type for agreeableness subscale scores. Similar to this result, Caruso, Witkiewitz, Belcourt-Dittloff, and Gottlieb (2001), Shields and Caruso (2004), and Yin and Fan (2000) found that sample type was a statistically significant predictor for reliability. In contrast to these studies, Hess, McNab, and Basoglu (2014), Thompson and Cook (2002), and Wallace and Wheeler (2002) found that the mean reliability estimates were invariant across different sample types. Additionally, Wallace and Wheeler (2002) examined whether language was related to reliability estimates. In their results there were no statistically significant differences for different languages. However, the results indicated that coefficient alpha estimates were affected by the language in Wheeler, Vassar, Worley, and Barnes's (2011) study.

Graham, Diebels, and Barnow (2011), Henson, Kogan, and Vacha-Haase (2001), Hess et al. (2014), and Nilsson et al. (2002) examined how the length of the test differentiates reliability coefficients in meta-analytic RG, and they concluded that as the length of the test increase, the reliability estimates increase. In addition to these results, Caruso (2000) concluded that the most important factor was the scale length for reliability coefficient. However, in another study, it was found that the mean reliability was predicted higher as the number of items in the scale decreased (Hanson, Curry, & Bandalos, 2002).

Sample size, which can be another possible source of variability for reliability estimation, was examined by some researchers (Hanson et al., 2002; Henson et al. 2001; Viswesvaran & Ones, 2000). Hanson et al. (2002) explored a correlation between reliability estimation and sample size or gender homogeneity. They reported that reliability estimates were related to the sample size of the client or therapist. The correlations between sample size and reliability estimates fluctuated in both direction and size for all subscales in Henson et al's. (2001) study. In contrast to these results, Viswesvaran and Ones (2000) found that there were no correlations between sample size and reliability estimations.

The other moderator variables, which are age of research participants, type of research design, testing conditions, gender, sexual orientation, ethnicity and marital status, and standard deviation of the subscale scores were also examined by researchers (Barnes et al., 2002; Capraro & Capraro, 2002; Graham et al., 2006; Graham et al., 2011; Vicent et al., 2019, and Wheeler et al., 2011). While the reliability estimates did not differ by the gender, sexual orientation, ethnicity and marital status in Graham et al.'s (2006) study, testing conditions, type of research design, aim of the study (psychometric or applied), and the standard deviation of the subscale scores were found as the sources of variability in reliability coefficients (Barnes et al. 2002; Capraro & Capraro, 2002; Vicent et al., 2019, and Wheeler et al., 2011). Also, the results of these studies showed that reliability estimates were sensitive to the age of the sample (Barnes et al., 2002; Graham et al., 2011; Vicent et al., 2019; Yin & Fan, 2000).

When the studies in the literature were examined, there were studies which analyzed the reliability generalization, and also whether the reliability coefficient differs according to variables such as test length (or the number of items), sample size, sample type, gender, reliability coefficient type, study language, race, marital status, age, etc. In the results of some of these studies, it was seen that variables such as number of items and sample type were found as sources of variability in reliability (e. g. Caruso, 2000; Caruso et al., 2001; Hanson et al., 2002; Henson et al., 2001; Hess et al., 2014; Nilsson et al., 2002; Shields & Caruso, 2004; Yin & Fan, 2000). In contrast, some studies concluded that these variables did not affect the reliability coefficient (e.g., Graham et al., 2011; Hess et al., 2014; Thompson & Cook, 2002; Wallace & Wheeler, 2002). As seen in previous studies, the reliability of the measures obtained with different scales can be affected by different variables. In this study, by examining these studies, a meta-analytic RG analysis was carried out for the Turkish sample. And similar to the literature, it was investigated how general the reliability coefficients are in different number of items, sample types, and fields of study and whether the reliability coefficients were affected by these variables. Within the scope of the study, it was aimed to analyze the reliability generalization of the long and short forms of the Oxford Happiness Scale (OHS) (Argyle, Martin, & Lu, 1995; Hills & Argyle, 2002). The reason for choosing this scale in the study was that the studies in the field of positive psychology have increased in recent years, and happiness is one of the concepts that are frequently researched in the field of positive psychology (Compton & Hoffman, 2019). Also, considering that the feeling of happiness has an effect on many aspects of individuals' lives, it is extremely important to measure the structure of happiness reliably. When both Turkish and non-Turkish literatures were examined, it was seen that the OHS is frequently used to measure happiness (e.g. Demir, 2020; Francis & Crea, 2018; Francis, Ok, & Robbins, 2017; Lin, Imani, Griffiths & Pakpour, 2020; Okur & Totan, 2019; Yıldırım & Sezer, 2020). Considering that the reliability values of the studies using the OHS in the Turkish literature have been in a wide range (.29 - .97), these differences should be investigated, and the reliability should be generalized for the Turkish sample. It is thought that RG studies can contribute as important sources of information for test administrators and researchers by Vacha-Haase, Henson, and Caruso (2002). In line with all this, it is important to bring this study into Turkish literature and the field of education and psychology.

METHOD

This meta-analysis study was performed according to the PRISMA (Liberati et al., 2009) guidelines. According to that, two authors searched the databases independently, identified the studies by screening the titles and the abstracts, removed the duplicates, and assessed the full-text articles for including in meta-analysis. This section includes data collection tools, sample, coding of study characteristics, and data analysis.

Data Collection Tool

The studies that were searched at the databases of Google Scholar, YOK (Higher Education Institution in Turkey) national thesis/dissertation center, EBSCOhost via Gazi University Central Library, and finally Aydin Adnan Menderes University Library's databases (e.g. BMJ, Dergipark, DOAJ, Clinical

Key, SAGE, Science Direct, Springer Link, Taylor & Francis etc.) and published between 2011 and 2020, and which used the long and short forms of the OHS were included in this review.

Oxford happiness scale

In the psychology, educational and social sciences, different measurement tools have been developed in order to measure happiness according to the increase in the studies on the concept of happiness. One of the most commonly used measurement tools in measuring happiness is the Oxford Happiness Inventory (OHI). This scale was developed by Argyle, Martin, and Crossland (1989) and Argyle et al. (1995).

OHI has been developed similarly to the format of the Beck Depression Inventory. The inventory has consisted of 29 personal well-being items by reversing 20 items from Beck Depression inventory and adding nine items that reflect different aspects of happiness. The cross-cultural comparison of OHI has been made by applying to students in Australia, Canada, and America (Francis, Brown, Lester, & Philipchalk, 1998). At the same time, it has been adapted for many different cultures such as Israel and China (Francis & Katz, 2000; Lu & Shih, 1997). However, since this inventory was developed by applying it to clinical patients, it was observed that individual responses were directed towards one of the two main items when administered to non-patients. The means for a substantial portion of items could be below the corresponding standard deviations. This showed that the responses could be distributed uniformly, and the items might not be able to fully contribute to the measurement of happiness. To overcome these situations, Hills and Argyle (2002) revised the inventory and constituted the OHS.

OHS consists of 29 items which are 6-point Likert-scale, and these points are within the range of *strongly agree-strongly disagree*. Half of the scale items are reversed. Thus, it is thought to decrease the possibility of individuals to respond harmoniously or biased. In addition, in the same study, an 8-item short form of OHS was developed for situations when setting was limited (Hills & Argyle, 2002).

The adaptation study of OHS to Turkish was conducted by Doğan and Sapmaz (2012). They examined the psychometric properties of the scale by implementing 491 university students. While the validity of OHS was investigated by criterion-related validity methods and exploratory and confirmatory factor analyses (EFA, CFA), the reliability was investigated by internal consistency, split-half, and composite reliability methods. Accordingly, the Cronbach's Alpha coefficient and composite reliability coefficient were found .91, and the reliability coefficient obtained by the split-half method was found .86.

When the validity studies were examined, as a result of the EFA, a single-factor structure was obtained, as it was in its original form. It was concluded that the single-factor structure of the scale was preserved with CFA. The findings revealed that the Turkish form of OHS showed similar psychometric features to its original form.

The short form of OHS, which consists of eight items, was adapted to Turkish by Doğan and Çötök (2011). They applied the scale to 532 university students and evaluated the psychometric properties via EFA and CFA, internal consistency, and test-retest methods. In the item analysis, item 4 was excluded from the scale because the item-total correlation value was less than .30. The reliability and validity analyses after this stage were made with the remaining seven items. Cronbach's Alpha coefficient calculated from the data obtained from 321 students were found as .74. In the test-retest reliability study, OHS-S was applied to 81 students at two-week intervals, and the correlation was found .85 between the two administrations.

The EFA was showed that the scale has a single-factor structure as its original form does. It was concluded that the single-factor structure of the scale was confirmed by CFA. As a result, it was determined that OHS-7 was a valid and reliable measurement tool to measure the happiness of Turkish students. In this review, studies administering the long or short form of the OHS, which was adapted to Turkish and was analyzed in terms of validity and reliability, were searched.

Search Strategy

The process for selecting studies was given in Figure 1 (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). When Figure 1 was examined, it was seen that the first stage was searching. The search of articles and thesis was carried out in the following databases: Google Academic, YOK national thesis/dissertation center, all databases in Gazi University Central Library, and Aydin Adnan Menderes University Library (Databases were presented in Appendix A). The used keywords were “oxford happiness” and “oxford mutluluk”. The search was carried out spanning the years 2011 to 2020 because the short and long forms of the OHS were adapted to Turkish in 2011 and 2020 respectively. When the specified databases were searched with keywords, it was seen that there were 6906 studies in total. First, the studies without Turkish sample groups were eliminated, and double coding was avoided. In addition, for studies involving more than one reliability coefficient, each coefficient was coded separately. Therefore, a total of 206 studies were coded from 6906 studies. Later, these 206 studies were examined according to inclusion criteria [i) They must be published in specified databases, ii) Cronbach’s Alpha reliability coefficients must be reported or can be calculable iii) They must include a sample group, sample size and scale form/number of item of study, iv) the sample group of study must consist of Turkish people and v) the language of the studies must be English or Turkish]. A summary of the phases of the meta-analysis was shown in Figure 1.

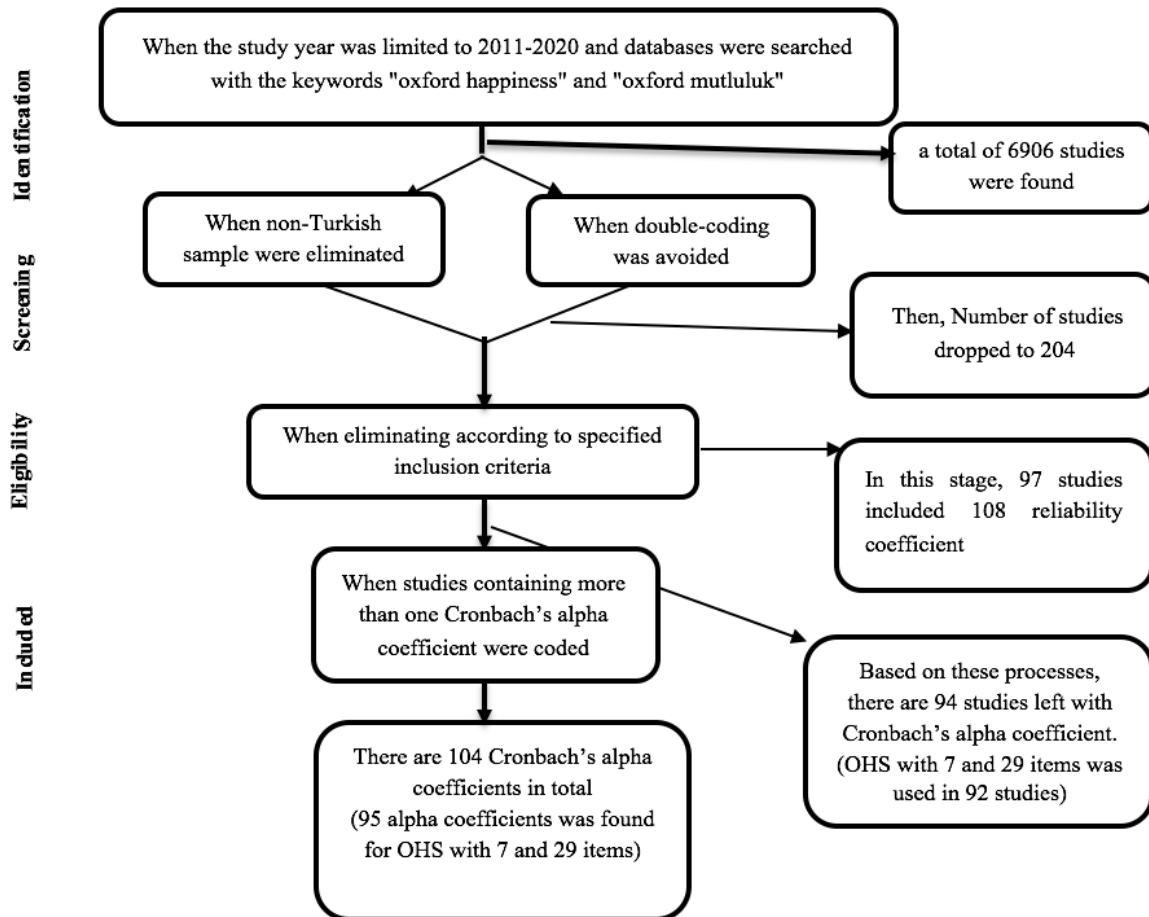


Figure 1. PRISMA Flowchart

After these phases of meta-analysis, the study group consisted of 92 studies of which 27 were thesis, and 65 were articles in accordance with the criteria determined by the researchers in this study. And 95 Cronbach’s Alpha coefficients were obtained from 92 studies that were presented in Appendix B. The selected studies were read and classified by two authors.

When Table 1, which has shown descriptive features of the studies included in the study, was examined, it was seen that seven studies were published between 2011-2015, and 85 studies were published between 2016-2020. Twenty studies were in English, and 72 studies were in Turkish. The short form was used in 56 of the studies and the long form of the scale in 36 of them. In addition, the sample type was coded as student for 50 studies and non-student for 42 studies. And, there were four studies with sample sizes ≤ 100 , 15 studies with sample sizes between 100 and 200, and 73 studies with sample sizes > 200 . Finally, the field of study was examined; it was observed that most studies (63) were in the field of social sciences. Also, it was seen that least studies (11) were in the field of sport sciences. Lastly, there were 18 articles or thesis for psychology/health sciences.

Table 1. Frequencies of the Studies According to Study Characteristics

Descriptive Variables	Categories	Number of Studies	Number of Cronbach's α
Number of items	7	56	58
	29	36	37
Type of Sample	Student	50	51
	Non-Student	42	44
Sample Size	≤ 100	4	6
	>100 and ≤ 200	15	16
	>200	73	73
Year of Study	2011-2015	7	7
	2016-2020	85	88
Language of Publication	Turkish	72	73
	English	20	22
Type of Publication	Article	65	68
	Thesis	27	27
Field of Study	Social Sciences (SS)	63	63
	Psychology (P)/Health Sciences (HS)	18	18
	Sport Sciences (SPS)	11	13
Total		92	95

Coding of Study Characteristics

After selecting the studies according to the inclusion criteria to the meta-analysis, the following sample and study characteristics were recorded by the researchers: (i)name of the article or thesis, (ii)name of the author(s) who conducted the study, (iii)year of the article or thesis, (iv)publication language of the study, (v)type of the study (article/thesis), (vi)type of the scale (the short form/the original form), (vii) reliability coefficient, (viii)type of reliability, (ix)sample size/the number of participants in the sample, (x)the number of items on the scale, (xi)fields of study and (xii)participant characteristics.

A total of 108 reliability coefficients were obtained from 97 studies. Of the 108, 104 were coefficient alpha; four coefficients were test-retest reliability, split-half reliability, and composite reliability estimates. However, the present study didn't characterize the scores by reliability type because of the small number of the reliability estimates differing from the coefficient alpha. Also, in some studies, it was observed that the item was removed or not used completely, and studies indicating a different number of items from the 7 and 29 items in the original scale forms were excluded from the study. Therefore, 92 studies remained when the studies that did not use all of the items were eliminated, and 95 alpha coefficients were obtained from these studies. Finally, 95 coefficient alpha values were analyzed for reliability generalization.

The inter-coder reliability was also examined for the data coded by the two authors according to the determined variables and criteria. The inter-coders reliability was calculated by the percent of agreement and Krippendorff's Alpha coefficient. For this, two coders coded for the same 10 studies and 11 reliability coefficients. These statistics were analyzed by SPSS 23 and SPSS macro that was developed by Hayes and Krippendorff (2007) and used for Krippendorff's Alpha coefficient. As a result of the analyses, the percent of agreement was .95, and the Krippendorff's alpha coefficient was

.94. These values were an indication that the inter-coder reliability is high. Krippendorff (2004) suggested that Krippendorff's Alpha coefficient should be at least .80, and he stated that $\alpha \geq .667$ is acceptable. Accordingly, inter-coder reliability is considered appropriate. Also, conflicts between the coders were examined by authors, and it has been determined that it was caused by the use of the keyboard. These conflicts that were detected were resolved.

Data Analysis

Reliability generalization studies provide reliability predictions to make a comparison between studies. In addition, it also examines the potential causes of variability in score reliability across studies (Graham et al., 2006). In this RG study, the generalizability of Cronbach's Alpha coefficients was investigated. Cronbach's Alpha is the square of the correlation because the reliability coefficients are variance-accounted statistics (Thompson & Vacha-Haase, 2000). Since the distribution of correlations isn't normal and has problematic standard errors, they must be transformed. Therefore, the raw alpha coefficients were transformed by *Fisher z-transformation*. Although Fisher's z-transformation was suggested for reliability coefficients calculated as Pearson correlation (e.g., test-retest, parallel forms) (Sánchez-Meca, López-López & López-Pina, 2013), recent studies have shown that Fisher z performed well and was very similar to other transformations in terms of empirical coverage probability (Romano, Kromrey, & Hibbard, 2010).

The random effects model (REM) which assumes that between-studies variance has been estimated greater than zero was used because of considering that the studies included in the research were obtained from different samples, fields, and years. Also, REM has been more realistic for real world applications (Field, 2003). In RG studies, there are a few heterogeneity estimators that are used for REM. Some of these estimators are Hunter-Schmidt, Hedges, DerSimonian and Laird, and the estimator based on maximum likelihood estimation (Maximum Likelihood-ML, Restricted ML-REML). In this study, the between-study variance, τ^2 , was estimated by DerSimonian and Laird.

The heterogeneity of Cronbach's Alphas was assessed by calculating the I^2 index as a function of Q statistic. The Q statistic was applied to test the assumption of homogeneity among the alpha coefficients. I^2 index is a possible measure of the amount of heterogeneity (Higgins & Thompson, 2002). It can be thought that I^2 values, which are approximately 25%, 50%, and 75%, reflect low, moderate, and large heterogeneity, respectively (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006).

To interpret the results, the mean effect sizes, their lower and upper confidence intervals obtained with Fisher z-transformation were back-transformed to the original metric of alpha coefficient. The predicted alpha coefficients were evaluated according to the .70 criterion level determined by Nunnally and Bernstein (1994). Values of .70 and above indicate that there is sufficient reliability for the internal consistency of the scale. The effect of the moderator variables on the variability of the reliability estimates was performed through Analog ANOVA. These moderator variables are type of scale (OHS, OHS-S), type of sample (student, non-student), and field of study (social sciences, psychology/health sciences, sport sciences). In addition, the variables of sample type and study field were analyzed as moderators separately for both OHS and OHS-S.

Lastly, publication bias was assessed by Egger's regression test (Egger, Smith, Schneider, & Minder, 1997), Begg and Mazumdar's rank correlation test (Begg & Mazumdar, 1994), Duval and Tweedie Trim and Fill (Duval & Tweedie, 2000a, 2000b) test for funnel plot asymmetry, fail-safe N method. Jamovi and Comprehensive Meta-Analysis V3 free trial (Retrieved from www.meta-analysis.com) was used for statistical analyses.

RESULTS

In this study, a meta-analysis of 95 Cronbach's Alpha coefficients was performed from moderator variables determined by examining literature. The distribution of alpha values in primary studies separately for each scale type is shown in Figure 2. Without the weighting factor, the average reliabilities of the alpha coefficients are .85 (SD = 0.08) and .74 (SD = 0.10) for OHS and OHS-S, respectively. The kurtosis and skewness coefficients are -2.50 (SE = 0.39), 6.73 (SE = 0.76) for OHS; and -1.95 (SE = 0.31), 9.13 (SE = 0.62) for OHS-S.

OHS Stem-and-Leaf Plot

Frequency Stem & Leaf
3.00 Extremes (= <.65)
1.00 7. 8
8.00 8. 01134444
18.00 8. 556677788889999999
7.00 9. 0001124

Stem width: .10
Each leaf: 1 case(s)

OHS-S Stem-and-Leaf Plot

Frequency Stem & Leaf
2.00 Extremes (= <.41)
1.00 6. 4
9.00 6. 788889999
19.00 7. 0000011222233334444
16.00 7. 5566667777888899
6.00 8. 111223
3.00 8. 556
2.00 Extremes (>=.94)

Stem width: .10
Each leaf: 1 case(s)

Figure 2. Distributions of Alpha Coefficients for OHS and OHS-S

Table 2 given below presents descriptive results for the estimates of alpha coefficients for general and moderator variables which are back-transformed to the alpha coefficient's original metric. Table 2 also shows 95% confidence interval for the estimated mean Cronbach's Alpha and the highest and lowest alpha values of the studies which constitute the RG meta-analysis.

As shown on the bottom line in Table 2, the reliability or the mean effect size of total OHS scores yielded a mean coefficient of .81 while the lower limit was .78 and the upper limit was .82 in 95% confidence interval. In addition, the reliability of total scores ranged from .29 to .97. Although there was a wide distribution of reliability estimates, the mean reliability estimate and limits of the confidence intervals are acceptable score reliabilities. For total reliability estimates, it can be said that they tend to be large and heterogeneous.

Table 2. Reliability Estimates of Oxford Happiness Scores across Studies for Different Moderator Variables

Category	<i>df</i>	Mean α (SD)	Z Value	95 % Confidence Interval				
				Lower	Upper	Min.	Max.	
Type of Scale								
OHS	36	0.87* (0.51)	35.98	0.85	0.88	0.50	0.94	
OHS-S	57	0.76* (0.50)	34.09	0.73	0.78	0.29	0.97	
Type of Sample								
Student	OHS	13	0.89* (0.24)	39.55	0.87	0.91	0.82	0.94
	OHS-S	36	0.75* (0.37)	25.42	0.71	0.78	0.29	0.97
Non-Student	OHS	22	0.85* (0.45)	30.03	0.83	0.87	0.54	0.92
	OHS-S	20	0.77* (0.34)	19.87	0.72	0.81	0.68	0.94
Field of Study								
Social Sciences (SS)	OHS	25	0.87* (0.33)	35.38	0.85	0.88	0.54	0.91
	OHS-S	38	0.77* (0.39)	25.54	0.73	0.80	0.29	0.97
Psychology (P)/Health Sciences (HS)	OHS	3	0.89* (0.10)	15.48	0.85	0.92	0.80	0.92
	OHS-S	12	0.74* (0.21)	13.91	0.67	0.79	0.41	0.85
Sport Sciences (SPS)	OHS	6	0.86* (0.38)	16.11	0.81	0.89	0.64	0.94
	OHS-S	5	0.72* (0.25)	8.67	0.61	0.81	0.68	0.74
Total	94	0.81* (0.72)		0.78	0.82	0.29	0.97	

Notes. Min.= minimum; Max.= maximum; OHS = Oxford Happiness Scale (original form which has consisted of 29 items); OHS-S= The short form of OHS (which has consisted of 7 items) and estimates use a random-effects model.

* $p < .05$

Table 2 also presents the mean alpha coefficients obtained for moderator variables. When the mean alpha coefficient was analyzed according to the type of scale, the mean alpha from the OHS-S was found .76 with a lower limit of .73 and an upper limit of .78 in 95% confidence interval. The mean effect size for the OHS was found .87 while the lower limit was .85, and the upper limit was .88 in 95% confidence interval. The reliability scores ranged between .50-.94 for the OHS and .29-.97 for OHS-S. The reliability scores range showed that especially OHS-S had lower coefficients than the OHS. The minimum reliability coefficients were below .70 for both types of scale (Nunnally & Bernstein, 1994). Again, the mean effect sizes and their 95% confidence interval limits were at an acceptable level for both types of scale. When the mean effect sizes were examined for two types of scale, the mean alpha coefficient for the OHS-S was smaller than OHS.

When Table 2 was examined according to characteristic of sample, the mean effect size estimates were higher in non-student sample for OHS-S. Despite that, the mean alpha value was higher in student sample for OHS. For OHS, the mean effect sizes were .89 and .85, respectively, in student sample and non-student sample. On the other hand, the mean effect sizes were found .75 and .77 respectively in student and non-student sample for OHS-S. For both types of sample there were wide distributions of reported alpha coefficients except OHS in student sample. The lowest reported alpha coefficient ($\alpha = .29$) was in student sample. And so, the minimum mean effect size was calculated as .75 (95%CI, .71-.78) in this sample.

With regard to field of study, it was seen that the mean alpha estimates were reported for three categories. The mean effect sizes obtained with the alpha coefficients of OHS were for Social Sciences ($\alpha = .87$, 95%CI, .85-.88), Psychology/Health Sciences ($\alpha = .89$, 95%CI, .85-.92), and Sport Sciences ($\alpha = .86$, 95%CI, .81-.89). Also, the mean effect sizes of OHS-S were .77 (95%CI, .73-.80) for Social Sciences, .74 (95%CI, .67-.79) for Psychology/Health Sciences, and .72 (95%CI, .61-.81) for Sport Sciences. According to these results, the mean alpha estimate for the field of Social Science was greater than the other fields for OHS-S. For OHS-S, the mean alpha estimates were almost close for all categories of field of study. For OHS, although the reliability estimates were close, the highest mean alpha value was in the field of psychology/health sciences.

In this study, the heterogeneity of Cronbach Alpha values was investigated. So, I^2 index for the amount of heterogeneity and Q test of homogeneity for the total scale were calculated. According to the results, the Q test was statistically significant with high heterogeneity coefficients. The estimates of Q for the

scale was $Q_{\text{Total}}(94) = 2639.66$, $p < .001$. The between-study variance, τ^2 , was estimated 0.08 by DerSimonian and Laird method. The I^2 index indicated that in 96.44%, the reliability coefficients had a large variability among the true effect size estimates. In the next step, the effect of the sub-group moderator variables was examined.

The Analog ANOVA was performed to examine the effect of the moderator variables on the variability of the reliability estimates. Whether the mean alpha coefficients differ according to the type of the scale was analyzed with the Analog ANOVA. The result was presented in Table 3.

Table 3. The Results of Analog ANOVA for Type of Scale

Moderator Variable	Categories	Q Statistics	df(Q)	I ²
Type of Scale	OHS	$Q_{\text{OHS}} = 415.63^*$	36	91.34%
	OHS-S	$Q_{\text{OHS-S}} = 1088.35^*$	57	94.76%
		$Q_{\text{within}} = 1503.98^*$	93	
		$Q_{\text{between(FEM)}} = 1135.69^*$	1	
		$Q_{\text{between(REM)}} = 50.75^*$	1	
		$Q_{\text{total}} = 2639.66^*$	94	

* $p < .05$

As shown in Table 3, Q_{total} of coefficient alpha values was found 2639.66 ($p = .00$), and it was statistically significant. Therefore, it can be said that the true variance estimate of reliability coefficients was statistically significant for all of the studies. In addition, it can be said that the variance within groups was statistically significant at the level of $p < .05$ since Q_{within} was 1503.98 ($p = .00$). When the difference between the groups was examined, it was seen that the $Q_{\text{between(REM)}}$ value was 50.75 ($p = .00$), and this value was significant. Accordingly, it can be said that the alpha coefficient was related to the scale type. When the variance in which the scale type explained for the alpha coefficient was examined, it was found that (1135.69/2639.66) .43 proportion of the true variance or 43.02% of the true variance was explained by the scale type. Based on this value, it can be said that the proportion of explaining the variance in the alpha coefficient of the scale type alone is high. When heterogeneity was examined for different scale types separately, heterogeneity was high in both scale types because Q statistics (Q_{OHS} and $Q_{\text{OHS-S}}$) was statistically significant at the level of $p < .05$, and I^2 were 94.76% and 91.34% for OHS-S and OHS, respectively. This may be due to lack of classification according to other variables ignored in this Q test. Some of these variables can be administration conditions, sample size, administration year, research type, etc. Due to the significant variance between the scale forms, it would be more meaningful to examine the effect of the moderator variables separately for OHS-S and OHS. As seen in Table 3, Q_{total} for OHS-S form data is also significant which means weighted sum of squares is much more than expected (df, k-1) by random, within study, variation. Forest plots for long form and short form were presented separately in Figure 3 and Figure 4.

When Figure 3 and Figure 4 were examined, it can be said that the reliability coefficients of all individual studies were statistically significant. In addition, it can be stated that the reliability coefficients were generally in the range of .80-1.00 and .60-1.00 for OHS and OHS-S, respectively.

Table 4. The Results of Analog ANOVA for Type of Sample in OHS-S

Moderator Variable	Categories	Q Statistics	df(Q)	I ²
Type of Sample	Student	$Q_{\text{student}} = 588.52^*$	37	93.88%
	Nonstudent	$Q_{\text{nonstudent}} = 394.30^*$	21	94.93%
		$Q_{\text{within}} = 982.82^*$	56	
		$Q_{\text{between(FEM)}} = 105.53^*$	1	
		$Q_{\text{between(REM)}} = 0.57$	1	
		$Q_{\text{total}} = 1088.35^*$	57	

* $p < .05$

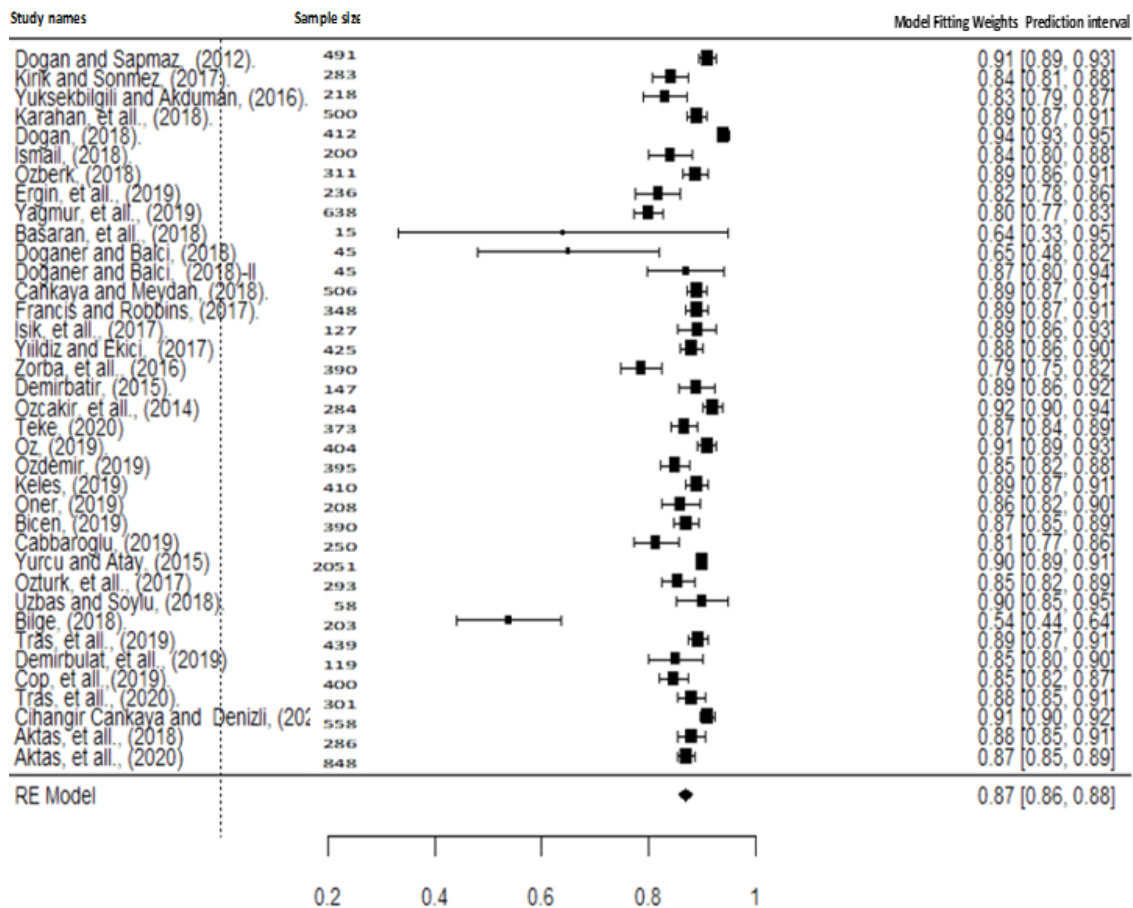


Figure 3. Forest Plot for OHS

The significance of the difference of alpha coefficients according to the type of the sample was analyzed by Analog ANOVA for OHS-S. When Table 4 was examined, it can be seen that the Q_{total} of coefficient alpha values was found 1088.35 ($p = .00$), and it was statistically significant. Therefore, it can be said that the variance was statistically significant for all of the studies that used OHS-S. In addition, it can be said that the variance within groups was statistically significant at the level of $p < .05$ since the Q_{within} was 982.82 ($p = .00$). When the difference between the groups was examined, it was seen that the $Q_{between(FEM)}$ value was 105.53 ($p < .05$). Accordingly, whether the sample consists of the students or not did have a statistically significant effect on the variability of alpha coefficient when FEM was used. In this case (105.53/1088.35), .10 proportion of variance or 10% of the true variance was explained by sample groups. However, this group difference could be overcome by using the REM analysis. As can be seen in Table 4, when the REM approach was used in the analysis, this variance was not significant anymore. Accordingly, whether the sample consists of students or not didn't have a statistically significant effect on the variability of alpha coefficients when REM was used. So, it can be said that the alpha coefficient was not related to the sample type of OHS-S for REM. When heterogeneity was examined for different sample types, it was high in both sample types for studies that used OHS-S. Because Q statistics ($Q_{student}$ and $Q_{nonstudent}$) was statistically significant at the level of $p < .05$, and I^2 were respectively 93.88% and 94.93% for student sample and non-student sample. This may be due to the lack of classification according to other variables ignored in this Q test. Some of these variables can be study field, administration conditions, sample size, administration year, research type, etc. Some further studies are needed to explain the remaining heterogeneity in OHS-S form reliability estimates.

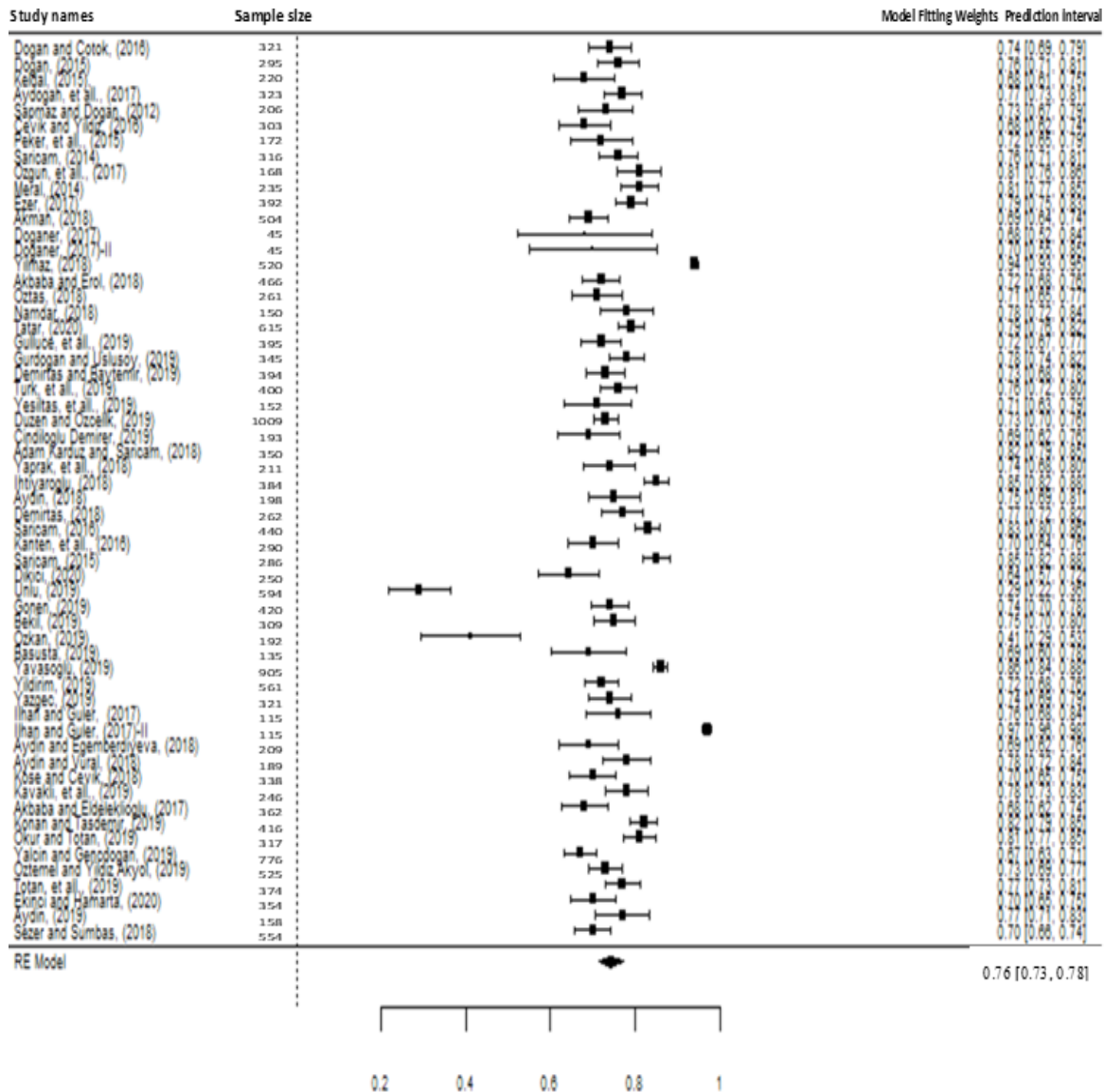


Figure 4. Forest Plot for OHS-S

Table 5. The Results of Analog ANOVA for Type of Sample in OHS

Moderator Variable	Categories	Q Statistics	df(Q)	I ²
Type of Sample	Student	Q _{student} = 76.61*	13	83.03%
	Nonstudent	Q _{nonstudent} = 292.70*	22	92.48%
		Q _{within} = 369.31*	35	
		Q _{between(FEM)} = 46.32*	1	
		Q _{between(REM)} = 7.90*	1	
		Q _{total} = 415.63*	36	

*p < .05

Table 5 presents the significance of the difference of alpha coefficients according to the type of sample for OHS. With regard to Table 5, Q_{total} of coefficient alpha values was found 415.63 (p = .00), and it was statistically significant. In addition, when we examined the variance within the groups, the studies separated by sample type were also heterogeneous in within groups. Q_{within} was 369.31, and p-value was 0.00 (p < .05). When the difference between the groups was examined, it was seen that the Q_{between(FEM)} value was 46.32 (p < .05). Accordingly, whether the sample consists of students or not did have a statistically significant effect on the variability of alpha coefficients when FEM was used. In

this case (46.32/415.63), .11 proportion of true variance or 11.14% of the true variance was explained by means of sample groups for OHS. Also, this group difference couldn't be overcome by using the REM analysis because, as can be seen in Table 5, when the REM approach was used in the analysis, $Q_{\text{between(REM)}}$ value was 7.90 ($p < .05$), and this value was significant. Accordingly, whether the sample consists of students or not had a still statistically significant effect on the variability of alpha coefficients when REM was used. Therefore, although the heterogeneity was significant in the within groups, it can be said that the mean alpha values of the studies separated according to the sample type differed significantly from each other. And it can be said that the alpha coefficient was related to the sample type for OHS. In addition, based on the proportion of true variance value, it can be said that the proportion of explaining the true variance in the alpha coefficient of the sample type alone is low. When heterogeneity was examined for different sample types, heterogeneity was high in both sample types. Because Q statistics (Q_{student} and $Q_{\text{nonstudent}}$) was statistically significant at the level of $p < .05$, and I^2 values were respectively 83.03% and 92.48% for sample of student and sample of nonstudent. This may be due to the lack of classification according to other variables ignored in this Q test. Some of these variables can be field of study, administration conditions, sample size, administration year, research type, etc. Some further studies are needed to explain the remaining heterogeneity in OHS long-form reliability estimates.

Table 6. The Results of Analog ANOVA for Field of Study in OHS-S

Moderator Variable	Categories	Q Statistics	df(Q)	I^2
Field of Study	Social Sciences	$Q_{\text{Social}} = 948.20^*$	38	95.99%
	Psychology/Health Sciences	$Q_{\text{Psychology/Health}} = 129.54^*$	12	90.74%
	Sport Sciences	$Q_{\text{Sport}} = 1.31$	5	0.00%
		$Q_{\text{within}} = 1079.06^*$	55	
		$Q_{\text{between(FEM)}} = 9.29^*$	2	
		$Q_{\text{between(REM)}} = 1.17$	2	
		$Q_{\text{total}} = 1088.35^*$	57	

* $p < .05$

The Analog ANOVA was performed to examine whether alpha coefficients showed a statistically significant difference according to field of study for OHS-S. As seen in Table 6, Q_{total} of coefficient alpha values was found 1088.35 ($p = .00$), and it was statistically significant. Therefore, it can be said that variance was statistically significant for all of the studies that used OHS-S. In addition, it can be said that the variance within groups was statistically significant at the level of $p < .05$ since Q_{within} was 1079.06 ($p = .00$). When the difference between the groups was examined, it was seen that the $Q_{\text{between(FEM)}}$ value was 9.29 ($p < .05$). Accordingly, whether the sample consists of students or not did have a statistically significant effect on the variability of alpha coefficients when FEM was used. In this case (9.29/1088.35), a .01 proportion of the true variance or 1.00% of the true variance was explained by field of study for OHS-S. However, this group difference can be overcome by using the REM analysis. As can be seen in Table 6, when the REM approach was used in the analysis, this variance was not significant anymore. Accordingly, whether the sample consists of students or not didn't have a statistically significant effect on the variability of alpha coefficients when REM was used. So, it can be said that the alpha coefficient wasn't related to the field of study for OHS-S for REM. When heterogeneity was examined for different fields of study, heterogeneity was high in social sciences and psychology/health sciences for studies that used OHS-S because Q statistics (Q_{Social} and $Q_{\text{Psychology/Health}}$) was statistically significant at the level of $p < .05$ and I^2 were respectively 95.99% and 90.74% for the field of social sciences and psychology/health sciences. As mentioned before, this may be due to the lack of classification according to other variables ignored in this Q test. Some of these variables can be sample type, administration conditions, sample size, administration year, research type, etc. Some further studies are needed to explain the remaining heterogeneity in OHS-S form reliability estimates. In addition, although heterogeneity was high in the fields of social sciences and psychology/health sciences, it was observed that there was low heterogeneity in the field of sports

sciences because $Q_{Sport} = 1.31$ ($p = 0.93$) was not statistically significant at the level of $p < .05$, and I^2 was 0.00% for field of sport sciences.

Table 7. The Results of Analog ANOVA for Field of Study in OHS

Moderator Variable	Categories	Q Statistics	df(Q)	I ²
Field of Study	Social Sciences	$Q_{Social} = 239.02^*$	25	89.54%
	Psychology/Health Sciences	$Q_{Psychology/Health} = 82.50^*$	3	96.36%
	Sport Sciences	$Q_{Sport} = 91.12^*$	6	93.42%
		$Q_{within} = 412.64^*$	34	
		$Q_{between(FEM)} = 2.99$	2	
		$Q_{between(REM)} = 1.83$	2	
		$Q_{total} = 415.63^*$	36	

* $p < .05$

The Analog ANOVA was performed to examine whether alpha coefficients showed a statistically significant difference according to field of study for OHS. As seen in Table 7, Q_{total} of coefficient alpha values was found 415.63 ($p = .00$), and it was statistically significant. Therefore, it can be said that the variance was statistically significant for all of the studies that used OHS. In addition, it can be said that the variance within groups was statistically significant at the level of $p < .05$ since Q_{within} was 412.64 ($p = .00$).

When the difference between the groups was examined, it was seen that the $Q_{between(FEM)}$ value was 2.99 ($p > .05$). Also, $Q_{between(REM)}$ value was 1.83 ($p = 0.40$), and this value wasn't statistically significant. Accordingly, whether the field of study is social sciences, psychology/health sciences, and sport sciences or not did not significantly affect the variability in alpha coefficients for both models. So, it can be said that the alpha coefficient was not related to the field of study for OHS. Already, (2.99/415.63) the 0.01 proportion of true variance or 1.00% of the true variance was explained by study fields for OHS. When heterogeneity was examined for different fields of study, it was high in social sciences, psychology/health sciences, and sport sciences for studies that used OHS-S since Q statistics (Q_{Social} , $Q_{Psychology/Health}$ and Q_{Sport}) was statistically significant at the level of $p < .05$, and I^2 values were respectively 89.54%, 96.36% and 93.42% for the fields of social sciences, psychology/health sciences, and sport sciences. This may be due to the lack of classification according to other variables ignored in this Q test. Some of these variables can be sample type, administration conditions, sample size, administration year, research type, etc. Some further studies are needed to explain the remaining heterogeneity in OHS long-form reliability estimates. Also, it can be said that the absence of a significant difference between these fields supports the high level of heterogeneity among the groups. As mentioned before, the scale type had a large variability source for the alpha coefficient. On the other hand, the moderator variables which were sample type and field of study weren't seen as the important sources of variability in the alpha coefficients. The reason for the high level of heterogeneity in the same fields of study, sample types, and scale types is that the studies come from different universes.

In this study, the meta-analysis was performed with only published articles and theses. Since the published studies generally have a high or significant effect size, taking only these studies into the meta-analysis may cause publication bias. Therefore, the publication bias was examined by rank correlation and regression test for funnel plot asymmetry and classic fail-safe N method. In the fail-safe N method, assuming the main effect of the studies to be added is zero, it is calculated how many studies are to be added so that the p-value isn't significant. And the calculated number's name is fail-safe N. If only a few studies are needed, there may be a concern that the effect is actually zero (Borenstein, Hedges, Higgins, & Rothstein, 2013). The fail-safe N was calculated as 4975 ($p < .01$). According to these results, it was seen that the number of studies to be added was quite high so that not the summary effect was significant. The other approach, funnel plot asymmetry is seen in Figure 4.

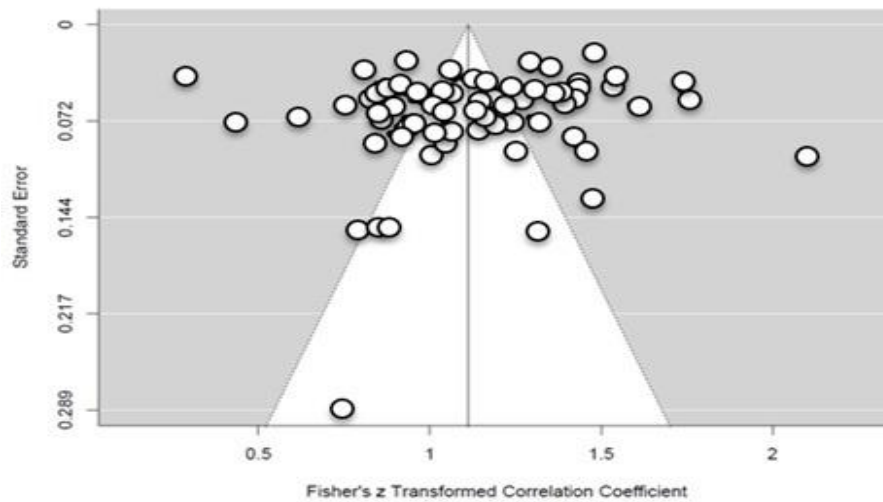


Figure 4. Funnel Plot for All Alpha Coefficients

In the funnel plot, studies are expected to be distributed symmetrically around the summary effect size. Although the studies were seen approximately symmetrically distributed to the right and left of the summary effect size, this interpretation is subjective (Borenstein et al., 2013). The rank correlation and regression tests were performed for more an objective interpretation. According to Egger's regression test, the regression intercept was not significant (intercept = -2.94, $p = .09$). The hypothesis was accepted to show that the regression constant didn't deviate from zero significantly. Begg and Mazumdar's rank correlation test also contributed to the lack of asymmetry in the funnel plot. According to that, Kendall's tau was not significant (Kendall's tau = -0.057, $p = .41$). It can be interpreted that there wasn't an asymmetry in the funnel diagram. In addition, according to Duval and Tweedie Trim and Fill test, there was no difference between the observed effect size and true effect size which was created to correct the effect caused by publication bias. As a result of the general symmetrical distribution of studies on both sides of the overall effect size, the difference was found zero. So the statistical tests for funnel plot asymmetry did not show any evidence of publication bias. Therefore, it can be said that all results were not likely to be the result of publication bias.

DISCUSSION and CONCLUSION

In this study, a meta-analytical reliability generalization analysis was conducted on OHS and OHS-S. In addition, it was investigated whether Cronbach's Alpha was affected by sample type, scale type, and study field. The results of this study showed that the mean Cronbach's Alpha coefficients obtained from both the OHS and OHS-S were at an acceptable level. The fact that these coefficients are high is an indication of the usability of the scale by both practitioners and researchers. When it was examined whether the reliability coefficient was affected by the scale type, a significant difference was observed between the two scale forms according to both REM and FEM. This difference was observed for the favor of OHS. Accordingly, it can be said that the measures obtained with the OHS, in general, are more reliable. In general, it is thought that more sensitive and more reliable measurements will be made as the number of items increases. And the results of this study support this idea. When the results of other studies were examined, it was seen that similar results were found. For example, Henson et al. (2001) and Nilsson et al. (2002) observed that reliability was higher for the long-form. Henson et al. (2001) stated that as the length of the test increase, the reliability estimates increase in all subscales except one. Also, Nilsson et al. (2002) found that the CDMSE long form's reliability coefficients were higher than the short form's. In contrast, Hanson et al. (2002), Hess et al. (2014), and Vacha-Haase (1998) observed that reliability was higher for the short form. Hanson et al. (2002) observed that the mean reliability coefficient obtained from the short form was slightly higher for both client and therapist versions of the Working Alliance Inventory. Hess et al. (2014) and Vacha-Haase (1998)

similarly concluded that the measures obtained from the short form were more reliable in a result of the RG analysis. As can be seen, while the effect of test length on reliability varies in the studies, it was observed that the reliability coefficient increases significantly as the test length increases for OHS in the Turkish sample. In the tests, it is recommended that the length of the test is as short as possible in terms of usability and it is long enough for acceptable reliability (McDonald, 1999). In this regard, the OHS is considered to be appropriate in terms of usability as it will not take much time to respond. Also, as a result of this study, it is obvious that the mean alpha coefficient of OHS for the Turkish sample is higher than OHS-S. In line with all of these, administering OHS instead of OHS-S may be more suitable for reliability for the Turkish sample. However, this situation may vary with the variance explained by the scale, the properties of the administration group, administration conditions, etc.

When the effects of the sample type on reliability were examined, it was seen that reliability was significantly different for students and non-students for the OHS-S according to FEM. But the true variance explained by sample groups was low for OHS-S. Although there was a significant effect in FEM, this group difference could be overcome by REM analysis. Therefore, researchers and practitioners may be advised to use REM analysis for such group differences. As a result, it was seen that the reliability wasn't significantly different for students and non-students for the OHS-S when REM was used. When the effects of the sample type on reliability were examined for OHS, it was seen that the reliability was significantly different for students and non-students for the OHS, according to both REM and FEM. Similar to these results, as it was in REM for OHS-S, Hess et al. (2014), Thompson and Cook (2002), Wallace and Weller (2002) found that reliability wasn't affected by the sample type. While Hess et al. (2014) separated sample types as student and professional, it was observed that Thompson and Cook (2002) distinguished as undergraduate, graduate, and faculty. And it was stated that there was no variability between the reliability coefficients of the groups in both studies. Also, Graham et al. (2011) concluded that the relationship between the proportion of college students in the sample and the reliability coefficient regarding the scores obtained with Locke–Wallace Marital Adjustment Test (LWMAT), Kansas Marital Satisfaction Scale (KMS), Quality of Marriage Index (QMI), and Marital Opinion Questionnaire (MOQ) wasn't significant. Contrary to these studies and similar to the results of OHS long-form Caruso et al. (2001), Vacha-Haase (1998), and Yin and Fan (2000) concluded that the sample type (student/non-student) affected reliability coefficients. As can be seen, while the effect of sample type on reliability varies in the studies, it was observed that the reliability coefficient didn't differ in student or non-student samples for OHS-S according to REM analysis in this study. The difference between the alpha coefficients is almost negligible, with about two per thousand for OHS-S. Also, the mean alpha coefficients were found to be high in both groups. These results are indicators of the availability of the OHS-S for both students and non-students for Turkish sample. In addition, while the effect of sample type on reliability varies in the studies, it was observed that the reliability coefficient differs for student or non-student samples for OHS in this study according to FEM and REM. This difference between the alpha coefficients was about four per thousand. But the proportion of explaining the true variance in the alpha coefficient of the sample type alone was quite low. Therefore, this difference wasn't at an important level. In addition, the mean alpha for student sample was higher than non-student sample, and the mean alpha coefficients were at an acceptable level for both OHS and OHS-S. The development of the OHS by applying it to students may be a factor in this. So, OHS is more suitable for students, but it can be used for both sample types. To summarize, it can be said to researchers and practitioners that both OHS and OHS-S can be used for both student groups and non-student groups for Turkish sample.

Finally, in the scope of the research, it was investigated how mean alpha was in different fields of study and whether the mean difference in reliability estimation was significant or not for these fields for both OHS and OHS-S. In OHS-S, the highest mean alpha was found in the field of social sciences, while the lowest mean alpha was found in the sport sciences. Also, the highest mean alpha was found in the field of psychology/health sciences, while the lowest mean alpha was found in the sport sciences for OHS. In line with all results, no major changes were observed in the reliability coefficients in all fields for all OHS forms according to REM. But in FEM analysis, the mean difference in reliability estimation was significant for OHS-S. When it was examined how much true variance was explained

with field of study, it was seen that explained variance was low and not important for OHS-S. Although there was a significant effect in FEM, this group difference could be overcome by REM analysis. Therefore, researchers and practitioners may be advised to use REM analysis for such group differences. As a result, it was seen that reliability wasn't significantly different for field of study for the OHS-S when REM was used. Based on this, it can be said that field of study is generally not effective in reliability estimation for all OHS forms. When the researches in the literature were investigated, it was seen that Vicent et al. (2019) analyzed the effect of study focus on reliability estimation by classifying study focus as applied and psychometric. And they concluded that the effect of study focus on reliability estimation for CAPS sub-dimensions was significant. However, they found that the variance in which the variability in reliability was explained by the study focus was low, and in meta-regression analysis, they found that it was the variable that least explained variance. It can be that the reason why this study is different from Vicent et al.'s (2019) research is that the study focus and field of study concepts are distinct each other and the classification is made differently. In such a case, the different measurement tools may have affected the differentiation of the results. Another study in the literature classified the study type as medical and nonmedical and research design as psychometric/others and experimental/others (Barnes et al., 2002). But they stated that they didn't examine the relationship between reliability and study type; they examined how much reliability was reported in journals in different contexts. Also, they founded that there were very low correlations between internal consistency coefficient and the contexts of psychometric/others or experimental/others. The low correlations found are similar to this study, but although research design and fields of study are similar, they are not the same. Consequently, reliability coefficients didn't differ significantly in different fields of study according to REM and were acceptable for all of them. Therefore, it is thought that the OHS's forms can be used in different fields.

The reliability estimates of OHS and OHS-S showed acceptable level in the present study. However, as mentioned above, as each measurement depends on the different conditions of the sample or settings, the results in this study are specific to these conditions. Therefore, it is necessary to calculate reliability based on their own data, besides the RG studies (Capraro & Capraro, 2002). To summarize in general, administering OHS instead of OHS-S may be more suitable for reliability for Turkish sample. Also, scale forms can be used for both student sample and non-student sample and can be used for each field of study. But, in generally, it is suggested that REM analysis should be performed for these variables since some group difference can be overcome when REM is used. Finally, in deciding which form of OHS to use, this situation may vary with the variance explained by the scale, the properties of administration group, administration conditions, etc.

The limitations of this study are transforming Cronbach's Alpha coefficients into Fisher Z scores, examining the variables of scale type, sample type, and field of study as sources of measurement error of reliability, and performing the analysis in the CMA program. Future investigations can examine and compare the reliability estimates which use other reliability estimators like Hakstian-Whalen (1976) and Bonett (2002) transform, etc. The RG studies can be made for other reliability estimates which address different sources of measurement errors. As the different sources of variability, study language, sample size, year of study, race, gender, age, marital status, mean and standard deviation of the measurements obtained from the scale, reliability type, research design, different sample type, etc. can be selected.

REFERENCES

- Aguayo, R., Vargas, C., Emilia, I., & Lozano, L. M. (2011). A meta-analytic reliability generalization study of the Maslach Burnout Inventory. *International Journal of Clinical and Health Psychology, 11*(2), 343-361. Retrieved from <http://www.redalyc.org/articulo.oa?id=33716996009>
- Anastasi, A. (1982) *Psychological testing* (5th ed.). New York: Macmillan.
- Argyle, M., Martin, M., & Crossland, J. (1989). Happiness as a function of personality and social encounters. In J. P. Forgas & J. M. Innes (Eds.), *Recent advances in social psychology: An international perspective* (pp. 189- 203). Amsterdam: North Holland, Elsevier Science.

- Argyle, M., Martin, M., & Lu, L. (1995). Testing for stress and happiness: The role of social and cognitive factors. In C. D. Spielberger, I. G. Sarason, J. M. T. Brebner, E. Greenglass, P. Laungani, & A. M. O'Roark (Eds.), *Series in stress and emotion: Anxiety, anger, and curiosity*, (Vol. 15). *Stress and emotion: Anxiety, anger, and curiosity* (p. 173-187). Washington, DC: Taylor & Francis.
- Armor, D. J. (1973). Theta reliability and factor scaling. *Sociological Methodology*, 5(1973-1974), 17-50. doi: 10.2307/270831
- Barnes, L. L., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger state-trait anxiety inventory. *Educational and Psychological Measurement*, 62(4), 603-618. doi: 10.1177/0013164402062004005
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50(4), 1088-1101. doi: 10.2307/2533446
- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, 62(4), 570-589. doi: 10.1177/0013164402062004003
- Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, 27(4), 335-340. doi: 10.3102/10769986027004335
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2013). *Introduction to meta-analysis*. UK: John Wiley & Sons.
- Bornmann, L., Mutz, R., & Daniel, H. D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PloS One*, 5(12), 1-10. doi: 10.1371/journal.pone.0014331
- Capraro, R. M., & Capraro, M. M. (2002). Myers-briggs type indicator score reliability across: Studies a meta-analytic reliability generalization study. *Educational and Psychological Measurement*, 62(4), 590-602. doi: 10.1177/0013164402062004004
- Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and Psychological Measurement*, 60(2), 236-254. doi: 10.1177/00131640021970484
- Caruso, J. C., Witkiewitz, K., Belcourt-Dittloff, A., & Gottlieb, J. D. (2001). Reliability of scores from the Eysenck Personality Questionnaire: A reliability generalization study. *Educational and Psychological Measurement*, 61(4), 675-689. doi: 10.1177/00131640121971437
- Compton, W. C., & Hoffman, E. (2019). *Positive psychology: The science of happiness and flourishing*. Thousands Oak, CA: SAGE Publications.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Ohio: Cengage Learning.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. doi: 10.1007/bf02310555
- Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement*, 25(2), 291-312. doi: 10.1177/001316446502500201
- Demir, Ü. (2020). Aile özellikleri ve mutluluk: Çanakkale'de lise öğrencileri üzerine bir araştırma. *Kastamonu Eğitim Dergisi*, 28(3), 1296-1306. Retrieved from <https://dergipark.org.tr/en/download/article-file/1109602>
- Doğan, T., & Çotok, N. A. (2011). Oxford mutluluk ölçeği kısa formunun Türkçe uyarlaması: Geçerlik ve güvenilirlik çalışması. *Türk Psikolojik Danışma ve Rehberlik Dergisi*, 4(36), 165-172. Retrieved from <http://www.turkpsdr.org/index.php/pdr/article/view/100/101>
- Doğan, T., & Sapmaz, F. (2012). Oxford mutluluk ölçeği Türkçe formunun psikometrik özelliklerinin üniversite öğrencilerinde incelenmesi. *Düşünen Adam Psikiyatri ve Nörolojik Bilimler Dergisi*, 25(4), 297-304. Retrieved from <https://dusunenadamdergisi.org/storage/upload/pdfs/1586247664-tr.pdf>
- Duval, S., & Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89-98. doi: 10.1080/01621459.2000.10473905
- Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463. doi: 10.1111/j.0006-341x.2000.00455.x
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109), 629-634. doi: 10.1136/bmj.315.7109.629
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement* (pp. 105-146). New York: Macmillan Publishing Co., Inc.; American Council on Education.
- Field, A. P. (2003). The problems in using fixed-effects models of meta-analysis on real-world data. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, 2(2), 105-124. doi: 10.1207/S15328031US0202_02

- Francis, L. J., & Crea, G. (2018). Happiness matters: Exploring the linkages between personality, personal happiness, and work-related psychological health among priests and sisters in Italy. *Pastoral Psychology*, 67(1), 17-32. doi: 10.1007/s11089-017-0791-z
- Francis, L. J., & Katz, Y. J. (2000). Internal consistency reliability and validity of the Hebrew translation of the Oxford Happiness Inventory. *Psychological Reports*, 87(1), 193-196. doi: 10.2466/pr0.2000.87.1.193
- Francis, L. J., Brown, L. B., Lester, D., & Philipchalk, R. (1998). Happiness as stable extraversion: A cross-cultural examination of the reliability and validity of the Oxford Happiness Inventory among students in the UK, USA, Australia, and Canada. *Personality and Individual Differences*, 24(2), 167-171. doi: 10.1016/S0191-8869(97)00170-0
- Francis, L. J., Ok, Ü., & Robbins, M. (2017). Religion and happiness: A study among university students in Turkey. *Journal of religion and health*, 56(4), 1335-1347. Retrieved from http://glyndwr.repositorytest.guildhe.ac.uk:8080/9182/1/Robbins_Religion_and_happiness.pdf
- Gilmer, J. S., & Feldt, L. S. (1983). Reliability estimation for a test with parts of unknown lengths. *Psychometrika*, 48(1), 99-111. doi: 10.1007/BF02314679
- Graham, J. M., Diebels, K. J., & Barnow, Z. B. (2011). The reliability of relationship satisfaction: A reliability generalization meta-analysis. *Journal of Family Psychology*, 25(1), 39-48. retrieved from <https://pdfs.semanticscholar.org/4bad/75bc14e345e7dabfc01888b50b5ef502305d.pdf>
- Graham, J. M., Liu, Y. J., & Jeziorski, J. L. (2006). The dyadic adjustment scale: A reliability generalization meta-analysis. *Journal of Marriage and Family*, 68(3), 701-717. doi:10.1111/j.1741-3737.2006.00284.x
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255-282. doi: 10.1007/bf02288892
- Hakstian, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika*, 41(2), 219-231. doi: 10.1007/BF02291840
- Hanson, W. E., Curry, K. T., & Bandalos, D. L. (2002). Reliability generalization of working alliance inventory scale scores. *Educational and Psychological Measurement*, 62(4), 659-673. doi: 10.1177/0013164402062004008
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77-89. doi: 10.1080/19312450709336664
- Heise, D. R., & Bohrnstedt, G. W. (1970). Validity, invalidity, and reliability. *Sociological Methodology*, 2(1970), 104-129. doi: 10.2307/270785
- Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the teacher efficacy scale and related instruments. *Educational and Psychological Measurement*, 61(3), 404-420. doi: 10.1177/00131640121971284
- Hess, T. J., McNab, A. L., & Basoglu, K. A. (2014). Reliability generalization of perceived ease of use, perceived usefulness, and behavioral intentions. *Mis Quarterly*, 38(1), 1-28. Retrieved from <https://www.jstor.org/stable/26554866>
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539-1558. doi: 10.1002/sim.1186
- Hills, P., & Argyle, M. (2002). The Oxford Happiness Questionnaire: A compact scale for the measurement of psychological well-being. *Personality and Individual Differences*, 33(7), 1073-1082. doi: 10.1016/S0191-8869(01)00213-6
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index? *Psychological Methods*, 11(2), 193-206. doi: 10.1037/1082-989X.11.2.193
- İlhan, M., & Güler, N. (2017). Likert tipi ölçeklerde olumsuz madde ve kategori sayısı sorunu: Rasch modeli ile bir inceleme. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(3), 321-343. doi: 10.21031/epod.321057
- Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA: Sage Publications.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.
- Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika*, 28(3), 221-238. doi: 10.1007/bf02289571
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160. doi: 10.1007/bf02288391
- Li, A., & Bagger, J. (2007). The Balanced inventory of desirable responding (BIDR): A reliability generalization study. *Educational and psychological measurement*, 67(3), 525-544. doi: 10.1177/0013164406292087

- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology*, 62(10), 1-34. doi: 10.1016/j.jclinepi.2009.06.006
- Lin, C. Y., Imani, V., Griffiths, M. D., & Pakpour, A. H. (2020). Psychometric properties of the Persian Generalized Trust Scale: Confirmatory factor analysis and Rasch models and relationship with quality of life, happiness, and depression. *International Journal of Mental Health and Addiction*. doi: 10.1007/s11469-020-00278-0
- Lu, L., & Shih, J. B. (1997). Personality and happiness: Is mental health a mediator? *Personality and Individual Differences*, 22(2), 249-256. doi: 10.1016/S0191-8869(96)00187-0
- McDonald, R. P. (1985). *Factor analysis and related methods*. London: Lawrence Erlbaum Associates Publishers.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. London: Lawrence Erlbaum Associates Publishers.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7), 1-6. doi: 10.1371/journal.pmed.1000097
- Nilsson, J. E., Schmidt, C. K., & Meek, W. D. (2002). Reliability generalization: An examination of the career decision-making self-efficacy scale. *Educational and Psychological Measurement*, 62(4), 647-658. doi: 10.1177/0013164402062004007
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Okur, S., & Totan, T. (2019). Psikolojik iyi oluşu değerlendiren Bradburn duyulanım dengesi ölçeğinin Türkçede incelenmesi. *Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 6(2), 1-12. Retrieved from <https://dergipark.org.tr/en/download/article-file/935147>
- Raju, N. S. (1977). A generalization of coefficient alpha. *Psychometrika*, 42(4), 549-565. doi: 10.1007/bf02295978
- Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, 14(1), 57-74. doi: 10.1207/s15327906mbr1401_4
- Romano, J. L., Kromrey, J. D., & Hibbard, S. T. (2010). A Monte Carlo study of eight confidence interval methods for coefficient alpha. *Educational and Psychological Measurement*, 70(3), 376-393. doi: 10.1177/0013164409355690
- Sánchez-Meca, J., López-López, J. A., & López-Pina, J. A. (2013). Some recommended statistical analytic practices when reliability generalization studies are conducted. *British Journal of Mathematical and Statistical Psychology*, 66(3), 402-425. doi: 10.1111/j.2044-8317.2012.02057.x
- Shields, A. L., & Caruso, J. C. (2003). Reliability generalization of the alcohol use disorders identification test. *Educational and Psychological Measurement*, 63(3), 404-413. doi: 10.1177/0013164403063003004
- Shields, A. L., & Caruso, J. C. (2004). A reliability induction and reliability generalization study of the CAGE questionnaire. *Educational and Psychological Measurement*, 64(2), 254-270. doi: 10.1177/0013164403261814
- Taşdibi-Ünlü, F. (2019). *Üniversite öğrencilerinde yaşam değeri, yaşamın anlamı ve sosyal iyi olmanın mutluluğu yordamadaki rolü* (Yüksek lisans tezi). Muğla Üniversitesi Eğitim Bilimleri Enstitüsü, Muğla. <http://tez2.yok.gov.tr/>
- Thompson, B., & Cook, C. (2002). Stability of the reliability of libqual+™ scores a reliability generalization meta-analysis study. *Educational and Psychological Measurement*, 62(4), 735-743. doi: 10.1177/0013164402062004013
- Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement*, 60(2), 174-195. doi: 10.1177/0013164400602002
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58(1), 6-20. doi: 10.1177/0013164498058001002
- Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development*, 44(3), 159-168. doi: 10.1177/0748175611409845
- Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement*, 62(4), 562-569. doi: 10.1177/0013164402062004002
- Vicent, M., Rubio-Aparicio, M., Sánchez-Meca, J., & González, C. (2019). A reliability generalization meta-analysis of the child and adolescent perfectionism scale. *Journal of Affective Disorders*, 245(2019), 533-544. doi: 10.1016/j.jad.2018.11.049

- Viswesvaran, C., & Ones, D. S. (2000). Measurement error in "Big Five Factors" personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement*, 60(2), 224-235. doi: 10.1177/00131640021970475
- Wallace, K. A., & Wheeler, A. J. (2002). Reliability generalization of the life satisfaction index. *Educational and Psychological Measurement*, 62(4), 674-684. doi: 10.1177/0013164402062004009
- Wheeler, D. L., Vassar, M., Worley, J. A., & Barnes, L. L. (2011). A reliability generalization meta-analysis of coefficient alpha for the Maslach Burnout Inventory. *Educational and Psychological Measurement*, 71(1), 231-244. doi: 10.1177/0013164410391579
- Yıldırım, O., & Sezer, Ö. (2020). The relationship between nomophobia and trait anxiety, basic psychological needs, happiness in adolescents. *Journal of Human Sciences*, 17(2), 535-547. Retrieved from <https://j-humansciences.com/ojs/index.php/IJHS/article/view/5917/3375>
- Yin, P., & Fan, X. (2000). Assessing the reliability of Beck Depression Inventory scores: Reliability generalization across studies. *Educational and Psychological Measurement*, 60(2), 201-223. doi: 10.1177/00131640021970466

Appendix A. Search Databases

- Google Scholar,
- YOK (Higher Education Institution in Turkey) national thesis/dissertation center,
- Gazi University Central Library
 - EBSCOhost
 - Academic Research Complete,
 - Applied Science & Business Periodicals Retrospective,
 - Applied Science & Technology Index Retrospective,
 - Art Index Retrospective,
 - BIR Entertainment,
 - Book Index with Reviews,
 - Business Source Complete,
 - CINAHL Plus,
 - Dentistry & Oral Sciences Source,
 - ebook Collection,
 - Education Index Retrospective,
 - E-Journals, ERIC,
 - European Views of the Americas,
 - GreenFILE,
 - Humanities & Social Sciences Index Retrospective,
 - Library-Information Science & Technology Abstracts,
 - MasterFILE Complete,
 - MasterFILE Reference ebook Collection,
 - MathSciNet,
 - MEDLINE,
 - Newspaper Source Plus,
 - Newswires,
 - OpenDissertations,
 - Regional Business News,
 - Social Sciences Index Retrospective,
 - SPORTDiscus,
 - Teacher Reference Center,
 - ULAKBIM Turkish National Databases,
 - Web News
- Aydin Adnan Menderes University Library
 - ASCE,
 - BMJ,
 - CabDirect,
 - Cambridge University Press,
 - Clinical Key,
 - DergiPark,
 - Cochrane Library,
 - DOAJ,
 - Emerald Premier,
 - GALE-Archives Unbound,
 - IEEE/IEE Electronic Library,
 - IGI Global,
 - JOVE,
 - JSTOR
 - Archive Journal Content,
 - Nature Journals All/Academic Journals,
 - OECD,
 - Oxford Journals Online,
 - OVID Journals,
 - Philosophy Documentation Center,
 - SAGE,
 - Science Direct,
 - Springer Link/Palgrave Macmillan Journals,
 - Taylor & Francis,
 - UptoDate,
 - Wiley Online Library

Appendix B. Studies in Meta-Analysis

- Adam-Karduz, F. F., & Sarıcam, H. (2018). The relationships between positivity, forgiveness, happiness and revenge. *Romanian Journal for Multidimensional Education/Revista Romaneasca pentru Educatie Multidimensionala*, 10(4), 1-22. doi: 10.18662/rrem/68
- Akbaba, A. Y., & Eldeleklioğlu, J. (2019). Adaptation of positive mental health scale into Turkish: A validity and reliability study. *Journal of Family Counseling and Education*, 4(1), 44-54. doi: 10.32568/jfce.569976
- Akbaba, T. P., & Erol, D. (2019). Üniversite öğrencilerinde romantik ilişkilerde akılcı olmayan inançlar ile mutluluk arasındaki ilişkinin incelenmesi. *International Journal of Current Approaches in Language, Education and Social Sciences*, 1(1), 32-44. Retrieved from <https://dergipark.org.tr/en/download/article-file/771960>
- Akman, E. (2018). *Sağlık profesyonellerinde akış deneyiminin öznel iyi oluş üzerine etkisi: Bir kamu ve özel hastane örneği* (Master's thesis). Marmara University, Sağlık Bilimleri Enstitüsü, İstanbul. Retrieved from <https://tez2.yok.gov.tr/>
- Aktaş, E., Şahin, N., & Gürbüz, E. C. (2018). Mutluluk ve tüketim arasındaki ilişki: Mersin Üniversitesi İİBF öğrencileri örneği. *BENĞİ Dünya Yörük-Türkmen Araştırmaları Dergisi*, 2018(2), 86-104. Retrieved from <https://dergipark.org.tr/en/pub/bengii/issue/52630/693084>
- Aktaş, E., Şahin, N., & Gürbüz, E. C. (2020). Tüketimin mutluluk üzerindeki etkisi: Çukurova Bölgesi örneği. *Bulletin of Economic Theory and Analysis*, 5(1), 21-40. doi: 10.25229/beta.563857
- Aydın, M. (2018). Genç yetişkinlerde mutluluğun özgünlük ve kişisel erdemler açısından incelenmesi. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*, 6(6), 1023-1030. doi: 10.18506/anemon.424118
- Aydın, M. (2019). Genç yetişkinlerde mutluluk, maneviyat ve kanaat. *The Journal of Social Science*, 3(6), 439-448. doi: 10.30520/tjsosci.571198
- Aydın, M., & Egemberdiyeva, A. (2018). Üniversite öğrencilerinin psikolojik sağlamlık düzeylerinin incelenmesi. *Türkiye Eğitim Dergisi*, 3(1), 37-53. Retrieved from <https://dergipark.org.tr/tr/download/article-file/496083>
- Aydın, M., & Vural, G. Z. (2018). Üniversite öğrencilerinin beden imgelerinin yaşam niteliklerine etkisi. *Eğitim Kuram ve Uygulama Araştırmaları Dergisi*, 4(3), 111-121. Retrieved from <http://ekuat.com/articles/universite-ogrencilerinin-beden-imgelerinin-yasam-niteliklerine-etkisi.pdf>
- Aydoğan, D., Özbay, Y., & Büyüköztürk, Ş. (2017). Özgünlük Ölçeği'nin uyarlanması ve özgünlük ile mutluluk arasındaki ilişkide maneviyatın aracı rolü. *The Journal of Happiness & Well-Being*, 5(1), 38-59. Retrieved from https://www.researchgate.net/profile/Didem_Aydogan2/publication/316034028
- Basaran, Z., Çalışkan, F., Çolak, S., & Erdal, R. (2018, Nisan). *Sportif rekreasyon etkinliklerinin yaşlıların mutluluk ve mental iyi oluş düzeylerine etkisi*. Congress Papers of the Association of Sports Sciences, Alanya, Antalya. Retrieved from <http://eds.a.ebscohost.com/eds/pdfviewer/pdfviewer?vid=4&sid=43845fc5-1a53-462f-8805-967ec779916f%40sdc-v-sessmgr02>
- Başusta, B. (2019). *Ergen mutluluğunun ebeveynlerin kişilik özellikleri ve evlilik çatışması açısından incelenmesi*. (Master's thesis). Fatih Sultan Mehmet Vakıf University, Lisansüstü Eğitim Enstitüsü, İstanbul. Retrieved from <https://tez2.yok.gov.tr/>
- Bekil, M. (2019). *Öğretmenlerde mutluluğun yordayıcıları olarak sosyal bağlılık, özgecilik ve sosyal empati*. (Master's thesis). Muğla Sıtkı Kocman University, Eğitim Bilimleri Enstitüsü, Muğla. Retrieved from <https://tez2.yok.gov.tr/>
- Bıçen, G. (2019). *İşgören mutluluk düzeylerinin iş tatmini ve iş performansı üzerine etkileri: Konaklama işletmelerinde bir inceleme* (Master's Thesis). Hacı Bayram Veli University, Lisansüstü Eğitim Enstitüsü, Ankara. Retrieved from <http://tez2.yok.gov.tr/>
- Bilge, H. (2018). Belediye çalışanlarının mutlulukları: Manisa büyükşehir belediyesi örneği. *Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, 33(2), 657-678. doi: 10.24988/deuıibf.2018332808
- Cabbaroğlu, M. (2019). *Sportif rekreasyon etkinliği olarak zumba ve pilatesin yaşam doyumu ve mutluluk üzerine etkisi (Muğla İli örneği)* (Master's Thesis). Muğla Sıtkı Kocman University, Sosyal Bilimler Enstitüsü, Muğla. Retrieved from <http://tez2.yok.gov.tr/>
- Cihangir-Çankaya, Z., & Denizli, S. (2020). An explanation of happiness with secure attachment, basic psychological needs and hope: the case of Turkish university students. *European Journal of Educational Research*, 9(1), 433-444. doi: 10.12973/eu-jer.9.1.433
- Cihangir-Çankaya, Z., & Meydan, B. (2018). Ergenlik döneminde mutluluk ve umut. *Electronic Journal of Social Sciences*, 17(65), 207-222. doi: 10.17755/esosder.316977

- Cindiloğlu-Demirer, M. (2019). Kişi örgüt uyumunun iş performansı üzerine etkisi: Mutluluğun aracılık rolü. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 33(1), 283-301. Retrieved from <https://hdl.handle.net/11491/5438>
- Cop, R., Topuz, Y. V., Eru, O., & Yüzüak, A. (2019). Effects of relational experience and subjective well being on customer satisfaction. *Third Sector Social Economic Review*, 54(1), 502-513. doi: 10.15659/3.sektor-sosyal-ekonomi.19.03.1102
- Çevik, G. B., & Yıldız, M. A. (2016). Pedagojik formasyon öğrencilerinde umutsuzluk ile mutluluk arasındaki ilişkide benlik saygısının aracılık rolü. *Dicle Üniversitesi, Eğitim Fakültesi Dergisi*, 27(2016), 96-107.
- Demirbatır, R. E. (2015). Relationships between psychological well-being, happiness, and educational satisfaction in a group of university music students. *Educational Research and Reviews*, 10(15), 2198-2206. doi: 10.5897/ERR2015.2375
- Demirtaş, A. S. (2018). Duygu düzenleme stratejileri ve benlik saygısının mutluluğu yordayıcılığı. *Electronic Turkish Studies*, 13(11), 487-503. doi: 10.7827/TurkishStudies.13465
- Demirtaş, A. S., & Baytemir, K. (2019). Warwick-Edinburgh mental iyi oluş ölçeği kısa formu'nun Türkçe'ye uyarlanması: Geçerlik ve güvenilirlik çalışması. *Electronic Journal of Social Sciences*, 18(70), 654-666. doi: 10.17755/esosder.432708
- Dikici, İ. (2020). *Serbest zamanlarını gençlik merkezlerinde değerlendiren üniversite öğrencilerinin serbest zaman doyum, yaşam doyum ve mutluluk düzeylerinin incelenmesi* (Master's thesis). Muğla Sıtkı Kocman University, Sosyal Bilimler Enstitüsü, Muğla. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Doğan, M. N. (2018). *Hukuk fakültesi öğrencilerinin rekreasyonel aktivitelerden elde ettikleri faydaların ve mutluluk düzeylerinin incelenmesi* (Master's thesis). Gazi University, Sağlık Bilimleri Enstitüsü, Ankara. Retrieved from <http://tez2.yok.gov.tr/>
- Doğan, T. (2015). Kısa psikolojik sağlık ölçeği'nin Türkçe uyarlanması: Geçerlik ve güvenilirlik çalışması. *The Journal of Happiness & Well-Being*, 3(1), 93-102. Retrieved from <https://www.tayfundogan.net/wp-content/uploads/2016/09/K%C4%B1saPsikolojikSağlamlikOlcegi.pdf>
- Doğan, T., & Çötök, N. A. (2016). Oxford mutluluk ölçeği kısa formunun Türkçe uyarlanması: Geçerlik ve güvenilirlik çalışması. *Turkish Psychological Counseling and Guidance Journal*, 4(36), 165-172. Retrieved from <http://www.turkpsdrergisi.com/index.php/pdr/article/view/100>
- Doğan, T., & Sapmaz, F. (2012). Oxford mutluluk ölçeği Türkçe formunun psikometrik özelliklerinin üniversite öğrencilerinde incelenmesi. *Düşünen Adam Psikiyatri ve Nörolojik Bilimler Dergisi*, 25, 297-304. doi: 10.5350/DAJPN2012250401
- Doğaner, S. (2017). *Düzenli egzersiz programının bireylerin stres, mutluluk ve serbest zaman doyum düzeylerine etkisi* (Doctoral dissertation). Ankara University, Sağlık Bilimleri Enstitüsü, Ankara. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Doğaner, S., & Balcı, V. (2018). Effect of regular physical activity on individuals' stress, happiness and leisure satisfaction levels. *Spormetre*, 16(3), 132-148. doi: 10.1501/Sporm_0000000382
- Düzen, A. Ç., & Özçelik, İ. Y. (2019, Kasım). *Investigation of psychological resilience and happiness levels of active and non-active high schoolers*. Congress Papers of the Association of Sports Sciences, Antalya. Retrieved from <http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=2&sid=5ab81a1d-5df4-47d9-866a-74eb94150602%40sessionmgr4007>
- Ekinci, N., & Hamarta, E. (2020). Meslek yüksekokulu öğrencilerinin azim ile mutluluk düzeylerinin incelenmesi. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 15(21), 125-141. doi: 10.26466/opus.569805
- Ergin, R., Ermiş, E., Erilli, A., & Ağaoğlu, S. A. (2019, Kasım). *Badminton sporcularının mutluluk ve sosyalleşme düzeylerinin incelenmesi*. Congress Papers of the Association of Sports Sciences, Antalya. Retrieved from <http://eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?vid=1&sid=347184b0-a666-4a52-813a-9c4ca9385dfd%40sessionmgr103>
- Ezer, H. İ. (2017). *Ergenlik döneminde yaygın kullanılan savunma mekanizmaları ile psikolojik sağlık ve mutluluk düzeyi arasındaki ilişki: Hatay il merkezi örneği* (Master's thesis). Çag University, Sosyal Bilimler Enstitüsü, Mersin. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Francis, L. J., Ok, Ü., & Robbins, M. (2017). Religion and happiness: A study among university students in Turkey. *Journal of Religion and Health*, 56(4), 1335-1347. doi: 10.1007/s10943-016-0189-8
- Gönen, M., (2019). *Antrenör-sporcu ilişkisinin sporcuların durumluk kaygı, öfke ve öznel iyi oluş düzeylerine etkisi: taekwondo ve korumalı futbol örneği*. (Doctoral dissertation). Gazi University, Sağlık Bilimleri Enstitüsü, Ankara. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Güdü-Demirbulat, Ö., Saatçi, G., & Avcıkurt, C. (2019). İşkoliklik ve bireysel mutluluk arasındaki ilişkinin turizm akademisyenleri açısından incelenmesi. *MANAS Sosyal Araştırmalar Dergisi*, 8(4), 3500-3315. doi: 10.33206/mjss.460701

- Güllüce, A. Ç., Kaygın, E., & Borekci, N. E. (2019). Üniversite öğrencilerinin nomofobi düzeyi ile öznel iyi olma durumları arasındaki ilişkinin belirlenmesi: Ardahan örneği. *Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 37(4), 651-673. doi: 10.17065/huniibf.482061
- Gürdoğan, E. P., & Uslusoy, E. C. (2019). The relationship between quality of work life and happiness in nurses: A sample from Turkey. *International Journal of Caring Sciences*, 12(3), 1364-1371. Retrieved from [http://www.internationaljournalofcaringsciences.org/docs/7_gurdogan_original_12_3\[20352\].pdf](http://www.internationaljournalofcaringsciences.org/docs/7_gurdogan_original_12_3[20352].pdf)
- Işık, Z., Çetinkaya, N., & Işık, M. F. (2017). Mutluluğun iş tatmini üzerindeki rolü: Erzurum ili Palandöken Kış Turizm Merkezinde yer alan konaklama işletmelerindeki kadın çalışanlar üzerine bir uygulama. *Journal of Graduate School of Social Sciences*, 21(2), 457-471. Retrieved from <https://dergipark.org.tr/en/download/article-file/474092>
- İhtiyaroğlu, N. (2018). Analyzing the Relationship between happiness, teachers' level of satisfaction with life and classroom management profiles. *Universal Journal of Educational Research*, 6(10), 2227-2237. Retrieved from <http://files.eric.ed.gov/fulltext/EJ1192735.pdf>
- İlhan, M., & Güler, N. (2017). Likert tipi ölçeklerde olumsuz madde ve kategori sayısı sorunu: Rasch modeli ile bir inceleme. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(3), 321-343. doi: 10.21031/epod.321057
- İsmail, M. (2018). *Mutluluğun yordayıcısı olarak cinsel benlik şeması ve sosyal görünüş kaygısı* (Master's thesis). İstanbul Ticaret University, Sosyal Bilimler Enstitüsü, İstanbul. Retrieved from <http://tez2.yok.gov.tr/>
- Kanten, P., Kanten, S., & Dünder, G. (2016). Ücret tatmininin ve işin özelliklerinin işe gömülmüşlük üzerindeki etkisinde mutluluğun rolü. *İşletme Araştırmaları Dergisi*, 8(3) 64-88.
- Karahan, G. (2018). *Örgütsel iletişimde işe ilişkin duyuşsal iyilik algısı ile öznel mutluluk arasındaki ilişkinin incelenmesi: Türkiye'de havayolu şirketlerinde kabin memurları üzerine bir alan araştırması* (Doctoral dissertation). Maltepe University, İstanbul. Retrieved from <http://tez2.yok.gov.tr/>
- Kavaklı, M., Kozan, H. İ. Ö., Kesici, Ş., & Ak, M. (2019). How can we feel happy? The examination of relationships among happiness, mindfulness and forgiveness. *Research on Education and Psychology*, 3(2), 198-208. Retrieved from https://www.researchgate.net/publication/338103622_How_Can_We_Feel_Happy_The_Examination_of_Relationships_Among_Happiness_Mindfulness_and_Forgiveness
- Keldal, G. (2015). Warwick-Edinburgh mental iyi oluş ölçeği'nin Türkçe formu: Geçerlik ve güvenilirlik çalışması. *The Journal of Happiness & Well-Being*, 3(1), 103-115. Retrieved from <https://toad.halileksi.net/sites/default/files/pdf/warwick-edinburgh-mental-iyi-olus-olcegi-toad.pdf>
- Keleş, A. (2019). *Din görevlerinde dindarlık ve mutluluk arasındaki ilişki (Sivas örneği)* (Master's thesis). Sivas Cumhuriyet University, Sosyal Bilimler Enstitüsü, Sivas. Retrieved from <http://tez2.yok.gov.tr/>
- Kırık, A. M., & Sönmez, M. (2017). İletişim ve mutluluk ilişkisinin incelenmesi. *İnif e-Dergi*, 2(1), 15-26. Retrieved from <https://dergipark.org.tr/en/pub/inifedergi/issue/27641/292862>
- Konan, N., & Taşdemir, A. (2019). Öğretmenlerin örgütsel ikiyüzlülük algıları ile mutluluk düzeyleri algıları arasındaki ilişki. *Scientific Educational Studies*, 3(2), 132-152. doi: 10.31798/ses.655939
- Köse, A., & Çevik, A. (2018). Happiness as a predictor of attitude towards teaching profession: Pedagogical formation case. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, 7(3), 853-873. doi: 10.14686/buefad.393207
- Meral, B. F. (2014). Kişisel iyi oluş indeksi-yetişkin Türkçe formunun psikometrik özellikleri. *The Journal of Happiness and Well-Being*, 2(2), 119-131. Retrieved from <https://www.journalofhappiness.net/frontend/articles/pdf/v02i02/3.pdf>
- Namdar, A. (2018). *Bir grup öğrencide umut, kaygı ve mutluluk arasındaki ilişki* (Master's thesis). Uskudar University, Sosyal Bilimler Enstitüsü, İstanbul. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Okur, S., & Totan, T. (2019). Psikolojik iyi oluşu değerlendiren Bradburn duygulanım dengesi ölçeğinin Türkçede incelenmesi. *Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 6(2), 1-12. Retrieved from <https://dergipark.org.tr/en/download/article-file/935147>
- Öner, C. (2019). *İş yaşamında mutluluk, iş yaşam dengesi ve işe adanma ilişkisi* (Master's thesis). Bahcesehir University, Sosyal Bilimler Enstitüsü, İstanbul. Retrieved from <http://tez2.yok.gov.tr/>
- Öz, T. (2019). *Çiftlerde depresyon, mutluluk ve psikolojik iyi oluş ile evlilik doyumu arası ilişkilerde evlilik süresinin aracı rolünün incelenmesi* (Master's thesis). Fatih Sultan Mehmet Vakıf University, Lisansüstü Eğitim Enstitüsü, İstanbul. Retrieved from <http://tez2.yok.gov.tr/>
- Özberk, F. (2018). *Annelerin eğitimleri çalışıp çalışmama durumuna göre lise öğrencilerinin duygusal zekâ mutluluk ve sosyal kaygı düzeylerinin karşılaştırılması* (Master's thesis). İstanbul Arel University, Sosyal Bilimler Enstitüsü, İstanbul. Retrieved from <http://tez2.yok.gov.tr/>
- Özçakır, A., Doğan, F. O., Çakır, Y. T., Bayram, N., & Bilgel, N. (2014). Subjective well-being among primary health care patients. *PloS one*, 9(12), 1-15. doi: 10.1371/journal.pone.0114496

- Özdemir, Ş. (2019). *Büyük ölçekli fitness merkezi kullanıcılarının tatmininin, ilgilenim ve mutluluk düzeylerine etkisinin belirlenmesi* (Master's thesis). Istanbul Okan University, Sosyal Bilimler Enstitüsü, İstanbul. Retrieved from <http://tez2.yok.gov.tr/>
- Özgün, A., Yaşartürk, F., Ayhan, B., & Bozkuş, T. (2017). Hentbolcuların spora özgü başarı motivasyonu ve mutluluk düzeyleri arasındaki ilişkinin incelenmesi. *Uluslararası Kültürel ve Sosyal Araştırmalar Dergisi (UKSAD)*, 3(Special Issue 2), 83-94. Retrieved from <https://dergipark.org.tr/en/pub/intjcss/issue/33182/369309>
- Özkan, A. (2019). *Üniversite öğrencilerinin sosyal uyum düzeyleri ile anlamlı yaşam çabası ve mutluluk arasındaki ilişkinin incelenmesi*. (Master's thesis). Ufuk University, Sosyal Bilimler Enstitüsü, Ankara. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Öztaş, İ. (2018). *Farklı kurumlarda çalışan memurların serbest zaman doyum ve mutluluk düzeylerinin belirlenmesi* (Kırıkkale ili örneği) (Master's thesis). Agri Ibrahim Cecen University, Sosyal Bilimler Enstitüsü, Ağrı. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Öztemel, K., & Yıldız-Akyol, E. (2019). The predictive role of happiness, social support, and future time orientation in career adaptability. *Journal of Career Development*, 20(1), 1-14. doi: 10.1177/0894845319840437
- Öztürk, L., Meral, İ. G., & Yılmaz, S. S. (2017). Lisans öğrencilerinin mutluluk ve dindarlık ilişkisi: Kırıkkale Üniversitesi örneği. *Akademik Yaklaşımlar Dergisi*, 8(1), 23-39. Retrieved from <https://dergipark.org.tr/tr/download/article-file/316231>
- Peker, A., Eroğlu, Y., & Özcan, N. (2015). Özel gereksinimli çocuğa sahip anneler ile tipik gelişim gösteren çocuğa sahip annelerin psikolojik sağlık, iyilik hali ve mutluluk düzeylerinin incelenmesi. *Sakarya University Journal of Education*, 5(3), 142-150. doi: Retrieved from <https://dergipark.org.tr/tr/download/article-file/192381>
- Sapmaz, F., & Doğan, T. (2012). Mutluluk ve yaşam doyumunun yordayıcısı olarak iyimserlik. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 8(3), 63-69. Retrieved from https://www.academia.edu/5249105/Mutluluk_ve_Ya%C5%9Fam_Doyumunun_Yorday%C4%B1c%C4%B1s%C4%B1_Olarak_%C4%B0yimserlik
- Sarıçam, H. (2014). Belirsizliğe tahammülsüzlüğün mutluluğa etkisi. *Sosyal Bilimler Dergisi*, 4(8), 1-12. Retrieved from <https://dergipark.org.tr/en/download/article-file/717400>
- Sarıçam, H. (2016). Examining the relationship between self-rumination and happiness: The mediating and moderating role of subjective vitality. *Universitas Psychologica*, 15(2), 383-396. doi: 10.11144/Javeriana.upsy15-2.errh
- Sarıçam, H. (2015). Metacognition and happiness: The mediating role of perceived stress. *Studia Psychologica*, 57(4), 271-283. Retrieved from https://www.researchgate.net/profile/Hakan_Saricam2/publication/286449970_
- Sezer, Ö., & Sumbas, E. (2018). Lise öğrencilerinin yaz tatillerini kullanım biçimlerinin bazı değişkenler açısından incelenmesi. *İnönü University International Journal of Social Sciences (INIJOSS)*, 7(1), 173-189. Retrieved from <http://dergipark.gov.tr/download/article-file/503986>
- Taşdibi-Ünlü, F. (2019). *Üniversite öğrencilerinde yaşam değeri, yaşamın anlamı ve sosyal iyi olmanın mutluluğu yordamadaki rolü* (Master's thesis). Muğla Sıtkı Kocman University, Eğitim Bilimleri Enstitüsü, Muğla. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Tatar, A. (2020). Madde yanıt kuramıyla A İş Stresi Ölçeği-20'nin geliştirilmesi: Güvenilirlik ve geçerlilik çalışması. *Anadolu Psikiyatri Dergisi*, 21(Ek Sayı 1), 14-22. doi: 10.5455/apd.77173
- Teke, M. (2020). *Yoksul kadınlarda mutluluk anlayışı* (Master's thesis). Sivas Cumhuriyet University, Eğitim Bilimleri Enstitüsü, Sivas. Retrieved from <http://tez2.yok.gov.tr/>
- Totan, T., Ercan, B., & Öztürk, E. (2019). Mutluluk ve benlik saygısının yalnızlıkla internet bağımlılığına etkilerinin incelenmesi. *EDU7*, 8(10), 20-35. Retrieved from <https://dergipark.org.tr/en/download/article-file/913720>
- Traş, Z., Öztemel, K., & Koçak, M. (2019). Öğretmen adaylarının mutluluk düzeylerinin bazı öznel niteliklerine göre incelenmesi. *Necmettin Erbakan Üniversitesi Ereğli Eğitim Fakültesi Dergisi*, 1(1), 47-56. Retrieved from <https://dergipark.org.tr/en/pub/neueefd/issue/45274/547448>
- Traş, Z., Öztemel, K., & Koçak, M. (2020). Üniversite öğrencilerinin mutluluk, yalnızlık ve sabır düzeyleri arasındaki ilişkinin incelenmesi. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 15(22), 878-894. doi: 10.26466/opus.575329
- Türk, R., Akkuş, Y., & Sönmez, T. (2019). Relationship between self-care ability and happiness in elderly individuals. *Cukurova Medical Journal*, 44(Suppl 1), 366-374. doi: 10.17826/cumj.560455
- Uz-Baş, A., & Soylu, Y. (2018). Pozitif duyguların psikolojik danışman adaylarının entelektüel becerileri ve mutluluk düzeylerine etkisi. *Journal of Higher Education & Science/Yükseköğretim ve Bilim Dergisi*, 8(2), 264-270. doi: 10.5961/jhes.2018.269

- Yağmur, Y., Oltuluoğlu, H., & Ergin, İ. O. (2019). İntrauterin dönemde fetal cinsiyetin annelerin mutluluk düzeyine etkisi. *ACU Sağlık Bil Derg*, 10(1), 89-93. doi: 10.31067/0.2018.98
- Yalçın, R. Ü., & Gençdoğan, B. (2019). Mutluluk saldırganlığın yordayıcısı mıdır? Üniversite öğrencileri ile bir yapısal eşitlik modeli çalışması. *Cumhuriyet International Journal of Education*, 8(3), 593-608. Retrieved from <https://dergipark.org.tr/tr/pub/cije/issue/48894/510491>
- Yaprak, P., Güçlü, M., & Ayyıldız Durhan, T. (2018). The happiness, hardiness, and humor styles of students with a bachelor's degree in sport sciences. *Behavioral Sciences*, 8(9), 1-21. doi: 10.3390/bs8090082
- Yavaşoğlu, E. (2019). *Eveli bireylerin mutluluk düzeyleri ile özgünlük ve değerler arasındaki yordayıcı ilişkiler* (Master's thesis). Istanbul Sabahattin Zaim University, Sosyal Bilimler Enstitüsü, İstanbul. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Yazgeç, G. (2019). *Doğa ve macera rekreasyonu etkinliklere katılan bireylerin serbest zaman doyum ve mutluluk düzeylerinin incelenmesi: Fethiye destinasyonu örneği* (Master's thesis). Manisa Celal Bayar University, Sosyal Bilimler Enstitüsü, Manisa. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Yeşiltaş, A., Şahin, S., & Serezli, G. (2019). Çalışan mutluluğunun ve işe bağlılığın örgüt performansına etkisi. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(30), 664-673. doi: 10.20875/makusobed.555411
- Yıldırım, O. (2019). *Ergenlerde akıllı telefondan yoksun kalma korkusu (Nomofobi) ile sosyodemografik değişkenler, temel psikolojik ihtiyaçlar, sürekli kaygı ve mutluluk arasındaki ilişkinin incelenmesi* (Master's thesis). Inonu University, Eğitim Bilimleri Enstitüsü, Malatya. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Yıldız, Y., & Ekici, S. (2017). Sporun üniversite öğrencileri üzerinde mutluluk ve sosyalleşme düzeylerine etkisinin incelenmesi. *Niğde Üniversitesi Beden Eğitimi ve Spor Bilimleri Dergisi*, 11(2), 181-187. Retrieved from https://www.researchgate.net/profile/Yasin_Yildiz6/publication/320755977_Investigation_Of_The_Effect_Of_Sports_On_The_Level_Of_Happiness_And_Socialization_Of_University_Students/links/5a6e1f75458515d407584ce6/Investigation-Of-The-Effect-Of-Sports-On-The-Level-Of-Happiness-And-Socialization-Of-University-Students.pdf
- Yılmaz, E. (2018). *Davranışsal ekonomide gelir ve mutluluk ilişkisi ve bir uygulama: Samsun ili örneği* (Master's thesis). KTO Karatay University, Sosyal Bilimler Enstitüsü, Konya. Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi>
- Yurcu, G., & Atay, H. (2015). Çalışanların öznel iyi oluşunu etkileyen demografik faktörlerin incelenmesi: Antalya ili konaklama işletmeleri örneği. *Manas Sosyal Araştırmalar Dergisi*, 4(2), 17-34. Retrieved from <https://dergipark.org.tr/en/pub/mjss/issue/40495/485114>
- Yüksekbilgili, Z., & Akduman, G. (2016). Bireysel mutluluk ve işkoliklik ilişkisi. *Kocaeli Üniversitesi Sosyal Bilimler Dergisi*, (31), 95-112. Retrieved from <https://dergipark.org.tr/tr/pub/kosbed/issue/25688/271103>
- Zorba, E., Pala, A., & Göksel, A. G. (2016). Examining the relation between emotional intelligence and happiness status of wellness trainers. *Journal of Education and Learning*, 5(3), 159-165. doi: 10.5539/jel.v5n3p159

Oxford Mutluluk Ölçeğinin Meta-Analizle Türk Örnekleminde Güvenirlilik Genellemesi

Giriş

Ölçme ve değerlendirmenin eğitim ve psikolojideki yeri ve önemi oldukça büyüktür. Buna göre, bu disiplinler arası ilişki ve dinamiğin sağlanması önemlidir. Ölçme ve değerlendirmenin en önemli iki kavramı ise *güvenirlilik* ve *geçerlilik*dir. Ölçme ve değerlendirmenin temel kuramlarından birisi olan klasik test kuramına göre güvenirlilik, ölçüm puanlarının tesadüfi hatalardan arınmışlık derecesi olarak tanımlanmaktadır. Ayrıca, güvenirlilik, aynı veya paralel testi alan bireylerin aldığı puanlar arasındaki tutarlılık anlamına da gelmektedir (Anastasi, 1982). Güvenirliliği yorumlayabilmek için geliştirilen formüller bulunmaktadır. Bu noktada, *güvenirlilik indeksi* ve *güvenirlilik katsayısı* kavramları arasındaki farkı belirtmek yararlı olacaktır. Güvenirlilik indeksi; gözlenen puanlar ile gerçek puanlar arasındaki ilişkiye odaklanırken, güvenirlilik katsayısı; paralel formlardan alınan puanlar arasındaki ilişkiyi ifade eder. İki kavram arasındaki matematiksel ilişkiye dayanarak, güvenirlilik katsayısının gerçek puan varyansının gözlenen puan varyansına oranı olduğu söylenebilir (Crocker & Algina, 2008). Güvenirlilik katsayısının hesaplanmasında araştırmacılar tarafından önerilen farklı formüller bulunmaktadır. Bu katsayılardan bazıları tek bir testin uygulamasını gerektirirken, bazıları birden fazla testin uygulamasını gerektirmektedir. Tek bir test uygulaması ile o teste ilişkin iç tutarlılık anlamındaki güvenirliliğin elde edilmesinde ve yorumlanmasında, alan yazında en sık kullanılan güvenirlilik katsayısı Cronbach Alfa'dır. Diğer güvenirlilik katsayıları gibi, Cronbach Alfa katsayısı da aynı ölçme aracının kullanıldığı farklı çalışmalarda, çalışmadan çalışmaya farklılık göstermektedir. Örneğin Oxford Mutluluk Ölçeği'nin kullanıldığı ve örneklemin üniversite öğrencilerinden seçildiği çalışmalardan birinde Cronbach Alfa katsayısı .29 iken (Taşdibi-Ünlü, 2019), başka bir çalışmada .97'dir (İlhan & Güler, 2017). Bir ölçme aracının güvenirlilik katsayısı bu örnekte olduğu gibi, farklı çalışmalardaki örneklem özelliklerine, örneklem büyüklüğüne, uygulama koşullarına, uygulama süresine vb. bağlı olarak değişebilmektedir. Güvenirlilik katsayılarında belirtilen çalışma özelliklerine bağlı olarak oluşan bu farklılıklar, güvenirliliğin genellemesini gerektirmiştir. Meta-analize dayalı olarak yapılan ilk güvenirlilik genellemesi çalışması Vacha-Haase (1998) tarafından yapılmıştır. Vacha-Haase'e göre, güvenirlilik genellemesi çalışmaları ile farklı ölçüm ve çalışmalardaki güvenirlilik katsayılarının kaynakları ve değişkenlik derecesi incelenebilir. Başka bir deyişle, güvenirlilik genellemesi çalışması, güvenirlilik katsayısının çalışmalar arasında farklılık gösterip göstermediğini inceleme olanağı tanımaktadır.

Alan yazında bu konuda yer alan çalışmalar incelendiğinde, genel etki büyüklüğünün kestirilmesinin yanı sıra güvenirlilik katsayısının; test uzunluğu (veya madde sayısı), örneklem büyüklüğü, örneklem türü, cinsiyet, yaş, ırk, güvenirlilik katsayısı türü vb. değişkenlere göre farklılık gösterip göstermediği incelenen çalışmalar da bulunmaktadır (Caruso, 2000; Caruso, Witkiewitz, Belcourt-Dittloff, & Gottlieb, 2001; Graham, Diebels, & Barnow, 2011; Hanson, Curry, & Bandalos, 2002; Wallace & Wheeler, 2002). Ülkemizde alanyazın incelendiğinde ise eğitim alanında meta-analiz çalışmalarının olması ile birlikte güvenirlilik katsayılarının meta-analiz yöntemi ile genelleştirildiği çalışmalara neredeyse hiç rastlanmamıştır. Bu çalışma kapsamında yabancı literatürdeki çalışmalar incelenerek, bu çalışmalara paralel bir şekilde, Oxford mutluluk ölçeğinin (OMÖ) kısa ve uzun formlarının meta-analiz yöntemi ile güvenirlilik genellemesinin incelenmesi amaçlanmıştır. Bu amaçla OMÖ kısa ve uzun formunun kullanıldığı çalışmaların meta-analizi ile Cronbach Alfa katsayılarının genel etki büyüklüğü kestirilmiştir. Bununla birlikte ölçek türü, çalışma alanı ve örneklem türü moderatör değişkenlerinin genel güvenirlilik kestirimini nasıl etkilediği araştırılmıştır.

Çalışmada OMÖ'nün seçilmesinin nedeni, pozitif psikoloji alanındaki çalışmaların son yıllarda artması ve mutluluğun pozitif psikoloji alanında en sık araştırılan kavramlardan biri olmasıdır (Compton & Hoffman, 2019). Ayrıca, hem Türk hem de yabancı literatür incelendiğinde, mutluluğun ölçülmesinde OMÖ'nün sıklıkla kullanıldığı görülmektedir (Demir, 2020; Francis & Crea, 2018; Lin, Imani, Griffiths, & Pakpour, 2020; Okur & Totan, 2019; Yıldırım & Sezer, 2020). Vacha-Haase, Henson ve Caruso (2002), güvenirlilik genelleme çalışmalarının test uygulayıcılar ve araştırmacılar için

önemli bilgi kaynağı olarak katkıda bulunabileceğini ifade etmişlerdir. Tüm bunlara paralel olarak bu çalışmayı ülkemiz alan yazınına ve eğitim ve psikoloji alanına kazandırmanın önemli olduğu düşünülmektedir.

Yöntem

OMÖ kısa ve uzun formunun genel güvenilirliğine ilişkin bilgi edinmeyi hedefleyen bu çalışmada meta-analiz Liberati ve diğerleri (2009) tarafından geliştirilen PRISMA yönergelerine uygun olarak yapılmıştır. Bu kapsamda, iki araştırmacı birbirinden bağımsız bir şekilde, Google Akademik, YÖK ulusal tez/tez merkezi, Gazi Üniversitesi Merkez Kütüphanesi ve Aydın Adnan Menderes Üniversitesi Kütüphanesi veri tabanlarında, 2011-2020 yılları arasında yayınlanan, Oxford mutluluk ölçeğinin uzun ya da kısa formunu kullanan çalışmaları taramıştır. Belirtilen veri tabanları “Oxford mutluluk” ve “Oxford happiness” anahtar kelimeleri ile taranarak, toplam 6906 çalışma; başlık ve özetlerine göre incelenmiştir. Ardından çift kodlama yapılan çalışmalar çıkarılmış ve tam metinler incelenmiştir. Daha sonra çalışmaya dahil etme kriterleri belirlenmiştir. Belirlenen ölçütler; i) belirlenen veri tabanlarında yayımlanmış olmak, ii) Cronbach Alfa katsayısının raporlanmış veya hesaplanabilir olması, iii) Örneklem grubunda, örneklem büyüklüğüne, ölçek formu ya da madde sayısına çalışmada yer verilmiş olması, iv) örneklem grubunun Türk bireylerden oluşmuş olması ve v) çalışma dilinin İngilizce ya da Türkçe olması şeklindedir. Son aşamada bu ölçütlere uygunluk kontrol edilerek 94 çalışma ve bu çalışmalarda da 104 Cronbach Alfa katsayısı olduğu tespit edilmiştir. Ancak bazı çalışmalarda madde atıldığı ya da tamamının kullanılmadığı gözlemlenmiş ve orijinal ölçek formlarındaki 7 ve 29 maddeden farklı sayıda madde sayısı belirten çalışmalar araştırmadan çıkarılmıştır. Sonuç olarak 27’si tez 65’si makale olmak üzere toplam 92 çalışma meta-analize dâhil olmuştur. Ayrıca, birden fazla güvenilirlik katsayısı içeren çalışmalar ayrı ayrı kodlandığı için toplam 95 Cronbach Alfa katsayısının meta-analizi gerçekleştirilmiştir. Güvenirlik katsayılarının dağılımını normalleştirmek için meta-analizden önce Alfa katsayılarına Fisher Z dönüşümü uygulanmıştır. Meta-analize dâhil olan çalışmalar farklı alanlarda, farklı yıllarda olduğu ya da farklı örneklemi içerdiği için etki büyüklüklerinin çalışmadan çalışmaya farklı olabileceği düşünülerek heterojenlik τ^2 ve Q istatistiğinin bir fonksiyonu olan I^2 istatistikleri ile incelenerek Rastgele etki modeli (REM) tercih edilmiştir. REM altında çalışmalar arası varyansın kestiriminde ise DerSimonian-Laird yöntemi kullanılmıştır.

Meta-analize dahil edilme kriterlerine göre seçilen çalışmaların kodlanması aşamasında belirtilen çalışma özellikleri ele alınmıştır: (i) çalışma adı, (ii) yazar(lar)ın adı, (iii) çalışmanın yayınlandığı yıl, (iv) çalışmanın yayın dili, (v) çalışma türü (makale/tez), (vi) ölçek türü (kısa form/uzun form) (vii) güvenilirlik katsayısı (viii) güvenilirlik türü, (ix) örneklem büyüklüğü, (x) ölçekteki madde sayısı, (xi) çalışma alanı ve (xii) katılımcı özellikleri. Belirtilen özelliklere göre çalışmalar iki araştırmacı tarafından kodlanmıştır ve kodlayıcılar arası uyum yüzdesi %94.80, Krippendorff Alfa katsayısı ise .94 bulunmuştur. Bu katsayı .80’den yüksek bulunduğu için kodlayıcılar arası güvenirliliğin yeterli düzeyde olduğu söylenebilir (Krippendorff, 2004). Yayın yanlılığının incelenmesinde huni diyagramı asimetrisi için Egger’in regresyon testi, Begg ve Mazumdar’ın sıra korelasyon testi ve Duval ve Tweedie’nin kırpma ve doldurma testi kullanılmıştır. Ayrıca fail-safe N yönteminden de yararlanılmıştır.

Araştırma kapsamında moderatör değişken olarak ölçek türü (kısa form/uzun form), örneklem türü (öğrenci/öğrenci değil) ve çalışma alanı (sosyal bilimler, psikoloji/sağlık bilimleri ve spor bilimleri) ele alınmıştır. Moderatör değişkenlere göre çalışmalar incelendiğinde, ölçek türü bağlamında 56 çalışma kısa formu, 36 çalışmanın uzun formu kullanmıştır. Örneklem türü bağlamında ise, 50 çalışmanın öğrenci örnekleme için veri topladığı, 42 çalışmanın ise öğrenci olmayan örneklemelerden veri topladığı görülmüştür. Son olarak çalışma alanları incelendiğinde ise sosyal bilimler alanında 63; spor bilimleri alanında 11; psikoloji/sağlık bilimleri kategorisinde ise 18 çalışma olduğu belirlenmiştir. Belirtilen bu moderatör değişkenlerin güvenilirlik kestiriminin değişkenliği üzerindeki etkisi, Analog ANOVA ile incelenmiştir.

Sonuç ve Tartışma

Araştırma sonucunda heterojenlik incelendiğinde, $\tau^2_{(DerSimonian-Laird)} = 0.08$, $Q_{total}(94) = 2639.66$, $p < .01$ ve $I^2 = \%96.44$ olduğu görülmüştür. Q istatistiğinin manidar olması ve I^2 'nin %75'ten yüksek olması güvenilirlik katsayılarının büyük oranda farklılaştığını ve heterojen dağıldıklarını göstermektedir. Yanlılığı yorumlama amacıyla, huni diyagramı için Egger'in regresyon testinin (sabit = -2.94, $p = .09$) ve Begg ve Mazumdar'ın sıra korelasyon testinin ($\tau = -0.06$, $p = .41$) manidar olmadığı bulunmuştur, bu nedenle huni diyagramının asimetrik olmadığı söylenebilir. Ayrıca, Duval ve Tweedie kırpma ve doldurma testine göre, gözlenen ve gerçek etki büyüklükleri arasında bir farklılık olmadığı görülmüştür. Huni diyagramının asimetrisine ilişkin bu testler sonucunda diyagramın simetrik bulunması yayın yanlılığının olmadığı bir göstergesidir. Ek olarak fail safe N sonuçları incelendiğinde N 4975 ($p < .01$) olarak hesaplanmıştır. Bu sonuç doğrultusunda yayın yanlılığının söz konusu olmadığı söylenebilir. Verilerin heterojen olarak dağılması nedeniyle REM'e dayalı olarak yapılan güvenilirlik genellemesine ve moderatör değişkenlerinin güvenilirlik katsayısına etkisine bu başlıkta yer verilmiştir.

Araştırma sonucunda tüm çalışmalarda yer alan Cronbach Alfa katsayılarının genel etki büyüklüğü .81 bulunmuştur. Bu katsayının alt ve üst limiti %95 güven aralığında .78-.82 olarak bulunmuştur. Buna dayalı olarak OMÖ ile elde edilen ölçümlerin güvenilirliğinin Türk örneklemini için genel itibarıyla yeterli düzeyde olduğu söylenebilir. Diğer yandan OMÖ'nün uzun ve kısa formuna ilişkin sonuçlar incelendiğinde, uzun forma ilişkin ortalama α 'nın .87 olduğu, kısa forma ilişkin ortalama α 'nın ise .76 olduğu görülmüştür. İki ölçek formunun ortalama α 'ları arasındaki fark Analog ANOVA ile incelendiğinde farkın $p < .05$ düzeyinde manidar olduğu gözlemlenmiştir. Bu farklılık uzun form lehinedir, buna dayalı olarak test uzunluğu arttıkça güvenilirlik katsayısının arttığı söylenebilir ve literatürde de benzer sonuçlara rastlanmıştır (Henson, Kogan, & Vacha-Haase, 2001; Nilsson, Schmidt, & Meek, 2002). Bu sonuçlara dayalı olarak, ortalama α 'ların her iki ölçek formunda da kabul edilebilir düzeyde bulunması nedeniyle Türkiye örnekleminde kullanılabilir olduğu düşünülmektedir. Ancak formların ortalama α 'ları arasındaki manidar farklılığa dayalı olarak kısa form yerine uzun form kullanımı daha uygun olabilir. Bu durum ölçeğin açıkladığı varyansa, uygulama grubunun özelliklerine, uygulama koşullarına vb. göre değişebilir. Uzun form ve kısa formun genel etki büyüklüğü arasında manidar bir farklılık olması nedeniyle örneklem büyüklüğünün ve çalışma alanının etkisi iki ölçek formu için ayrı ayrı incelenmiştir.

Moderatör değişkenlerden örneklem türü OMÖ kısa formu için incelendiğinde, örneklemini öğrencilerden oluşan ve öğrencilerden oluşmayan çalışmalar için ortalama α değerleri sırasıyla .75 ve .77'dir. İki örneklem türünün ortalama α 'ları arasındaki fark Analog ANOVA ile incelendiğinde; sabit etkiler modeline (SEM) göre manidar farklılık ($p < .05$) bulunmuştur. Grup içi heterojenliğin de yüksek olduğu kategorilerde REM'e göre ise manidar farklılık ($p < .05$) olmadığı gözlenmiştir. OMÖ uzun formu için sonuçlar incelendiğinde ise, örneklemini öğrencilerden oluşan ve öğrencilerden oluşmayan çalışmalar için ortalama α değerleri sırasıyla .89 ve .85 bulunmuş ve bu iki örneklem türünün ortalama α 'ları arasındaki fark ise her iki modele göre $p < .05$ düzeyinde manidar bulunmuştur. OMÖ uzun formu için örneklem türü, ortalama α 'nın değişkenliğini manidar olarak etkilese de örneklem türünün açıkladığı varyans %11.40 olup ortalama α 'nın değişkenliğini düşük bir oranda açıkladığı söylenebilir. Bu doğrultuda örneklem türünün güvenilirlik katsayısı kestirimi üzerinde etkisi ölçeğin kısa ve uzun formu için farklıdır. Graham ve diğerleri (2011), Thompson ve Cook (2002) ve Wallace ve Weller (2002) de OMÖ kısa formunda olduğu gibi örneklem türünün genel güvenilirlik kestirimine bir etkisi olmadığı sonucuna varmışlardır. Buna karşın, OMÖ uzun formundaki sonuçlara benzer şekilde Caruso ve diğerleri (2001), Vacha-Haase (1998) ve Yin ve Fan (2000) örneklem türünün genel güvenilirlik kestirimini etkilediği sonucuna varmışlardır. Bu sonuçlara dayalı olarak, OMÖ uzun formunun öğrenci örnekleminde daha güvenilir sonuçlar verdiği görülse de açıkladığı varyans oranı düşüktür, böylece iki ölçek formunun da Türkiye örneklemini için hem öğrenci hem de öğrenci olmayan örneklemelerde kullanılabilir olduğu söylenebilir.

Son olarak moderatör değişkenlerden çalışma alanı OMÖ kısa formu için incelendiğinde, ortalama α 'nın sosyal bilimler alanı için .77, psikoloji/sağlık bilimleri alanı için .74, spor bilimleri için .72

olduğu görülmüştür. Bu alanlara ait ortalama α 'lar arasındaki fark kısa form için Analog ANOVA ile incelendiğinde SEM'e göre manidar farklılık ($p < .05$) bulunsa da çalışma alanının gerçek varyansı açıklama oranı %1'den düşüktür. Rastgele etkiler modeli dikkate alındığında $p < .05$ düzeyinde manidarlık olmadığı gözlemlenmiştir. OMÖ uzun formu için sonuçlar incelendiğinde ise, ortalama α , sosyal bilimler alanı için .87, psikoloji/sağlık bilimleri alanı için .89, spor bilimleri için .86'tir. OMÖ uzun formu için bu alanlara ait ortalama α 'lar arasındaki fark ise her iki modele göre $p < .05$ düzeyinde manidar bulunmamıştır. Bu doğrultuda her iki form türü için de ortalama güvenilirlik katsayılarının yeterli düzeyde olması ve alanlar arasında manidar bir farklılık olmaması nedeniyle OMÖ formlarının Türkiye örnekleminde bu alanlar için uygun olduğu söylenebilir.

Sonuç olarak, ortalama α 'ların tüm moderatör değişken kategorilerinde kabul edilebilir düzeyde olması nedeniyle OMÖ'nün uygulayıcılar ve araştırmacılar açısından kullanılabilir olduğu belirtilebilir. Bu değişkenlere ait grup farklılıklarının üstesinden gelebilmek için REM ile analiz yapılması önerilebilir. Ölçek formlarından uzun formun ortalama güvenilirlik katsayısının kısa forma göre manidar bir şekilde yüksek olması nedeniyle kısa form yerine uzun formun kullanılması ölçümlerin güvenilirliği bağlamında tercih edilebilir. Ancak bu durum araştırmanın amacı, ölçeklerin açıkladığı varyans ve uygulama koşullarına bağlı olarak uygulayıcı ve araştırmacılar açısından değişkenlik gösterebilir.

Bu araştırmada OMÖ'nün meta-analitik güvenilirlik genellemesi yapılmış ve moderatör değişkenler ölçek türü, örneklem türü ve çalışma alanı olarak belirlenmiştir. Diğer araştırmacılar farklı ölçekler için güvenilirlik genellemesi çalışması yürütebilir ya da amaçları doğrultusunda farklı moderatör değişkenleri ele alabilirler. Farklı moderatör değişken olarak çalışma dili, örneklem büyüklüğü, çalışma yılı, ırk, cinsiyet, yaş, ölçekten elde edilen ölçümlerin ortalaması ve standart sapması, güvenilirlik türü, araştırma deseni, farklı örneklem türü vb. seçilebilir. Ayrıca bu çalışmada Türkiye örneklemini ele alınmış olup başka araştırmacılar farklı ülkelerin örneklemlerini ya da dünya genelinde yayınlanan ölçeklerle ilgili tüm çalışmaları ele alabilirler.