International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 476–488
https://doi.org/10.21449/ijate.1457753
journal homepage: https://dergipark.org.tr/en/pub/ijate
Research Article

Assessing the reliability of the Beck Depression Inventory adapted for Turkish culture: A meta-analysis

Yasemin Duygu Esen[1*], **Filiz İsnaç**[2], **H. Deniz Gülleroğlu**[3]

[1]Ministry of National Education, Eskisehir, Türkiye
[2]Ministry of National Education, Mersin, Türkiye
[3]Ankara University, Faculty of Education, Department of Educational Sciences, Ankara, Türkiye

ARTICLE HISTORY
Received: Mar. 28, 2024
Accepted: Nov. 25, 2024

Keywords:
Meta-analysis,
Beck Depression Scale,
Validity,
Reliability.

Abstract: This study aims to determine the reliability of the Beck Depression Inventory (BDI), a widely used tool for diagnosing depression—a condition that significantly impacts individuals' lives—through meta-analysis, an advanced statistical technique. To achieve this objective, studies conducted in Turkey between 1961 and 2021 that utilized the Beck Depression Inventory were included. Relevant studies were identified through searches in the "Higher Education Council (HEC) National Thesis Center" and Google Scholar databases. The inclusion criteria focused on studies that reported validity and reliability analyses and contained the keywords “Beck Depression Inventory” or “Beck Depression Scale.” The effect size was calculated using the Cohen's d formula. The meta-analysis revealed an overall reliability coefficient of .88 for the Beck Depression Inventory, which, based on Cohen's d classification, represents a large effect size.

1. INTRODUCTION

Scientific progress is cumulative, with each researcher building upon the knowledge and findings of their predecessors. The systematic accumulation and replication of empirical research underpin the steady advancement of scientific knowledge. However, scientists often face limitations in storing, organizing, and synthesizing empirical findings, which can impede this progress. Consequently, repeated studies in a particular field or on a specific topic often yield varying results. To better understand the reasons for these discrepancies and to evaluate the overall findings, a comprehensive summary or review of the studies is essential. This is where meta-analysis plays a critical role. Meta-analysis, a quantitative research method, can be defined as the process of combining the results of multiple studies into a single, comprehensive outcome (Borenstein *et al*., 2009). Arthur *et al*. (2001) describe meta-analysis as a set of statistical procedures used to derive general conclusions from primary sources and synthesize the results of multiple primary studies. Similarly, Dinçer (2014) defines meta-analysis as grouping studies with similar subjects, themes, or fields under specific criteria and interpreting their quantitative findings collectively. Meta-analysis offers several advantages, including integrating findings from independent studies, enhancing the validity of results, generalizing findings from studies with small sample sizes to larger populations, assessing whether

*CONTACT: Yasemin Duygu ESEN yaseminduyguesen@gmail.com Ministry of National Education, Eskisehir, Türkiye

e-ISSN: 2148-7456

variations between studies are due to chance, identifying the sources of variation, and resolving contentious issues (Glass, 1976; Hedges & Olkin, 1985).

Meta-analysis, widely utilized in the field of health, has increasingly gained significance in the field of education (Demiray, 2013; McDermott & Ebmeiern, 2009). It can be applied to various types of quantitative studies, including controlled clinical trials, quasi-experimental studies, and observational studies. While most meta-analyses focus on studies that aim to explain causal relationships, they can also be conducted on non-causal survey studies as well as on validity and reliability studies of diagnostic methods (Abramson, 1994).

Depression is a significant illness that profoundly impacts an individual's quality of life, engagement with life, and both professional and family dynamics (Belsher & Costello, 1988; Burcusa & Iacono, 2007; Monroe & Harkness, 2012; Solomon, 2000; Zis & Goodwin, 1979). If left undiagnosed and untreated, it can cause direct or indirect harm, including increasing the likelihood of alcohol and drug use and acting as a trigger for various medical conditions. The prevalence of suicide due to depression continues to rise, making it one of the leading causes of death. While depression is treatable, accurate diagnosis is a crucial first step. This relies on the validity and reliability of the measurement tools used for assessment. The Beck Depression Inventory (BDI) is a self-report scale developed by Beck in 1961, widely utilized for diagnosing depression and comprehensively evaluating the severity of depressive symptoms. According to the American Psychiatric Association (2000), the BDI is officially recognized for determining the initial diagnostic level of depression. It is one of the most frequently used tools in clinical studies as well as in health and education research due to several advantages. These include its applicability to both healthy individuals and psychiatric patients, its ease of use without requiring special training, its concise format, straightforward scoring and interpretation, and its well-established validity and reliability.

Due to its numerous advantages, the Beck Depression Inventory (BDI) has been adapted for use in various cultures and is widely utilized globally. The literature includes numerous studies examining the psychometric properties of the scale across diverse cultural contexts and populations, consistently demonstrating its high reliability (García-Batista *et al*., 2018; Hamidi *et al*., 2015). For instance, Wiebe *et al*. (2005) conducted a comprehensive analysis comparing the original English version of the BDI with its Spanish adaptation. The original version of the scale reported a Cronbach's alpha coefficient of .89, while the Spanish adaptation demonstrated a slightly higher coefficient of .91, reflecting high reliability in both versions. Additionally, the test-retest reliability method confirmed the scale's reliability at an acceptable level.

In their extensive review, Wang and Gorenstein (2013) analyzed approximately 118 articles that assessed the BDI's psychometric properties. Their findings revealed that the Cronbach's alpha coefficient reported in these studies ranged from .84 to .94, while test-retest reliability values varied from .73 to .96. This study concluded that the BDI is a reliable tool for diagnosing depression. A broader review of the international literature further supports these findings, indicating that the scale's reliability values are consistently high and comparable across different cultural adaptations. The Beck Depression Inventory (BDI) was adapted to Turkish culture by Şahin in 1988, with its reliability initially evaluated by Hisli using various methods. The split-half reliability coefficient was reported as .74, and the Cronbach's alpha coefficient was determined to be .80. However, reliability values for the BDI in studies conducted within Turkish samples vary widely, ranging from .53 to .98. For example, Orhan (2019) reported a reliability coefficient of .88, while Alver and Uğuryol (2016) noted a reliability value of .98. This considerable variation in reliability values, differing significantly from those reported in other cultures, forms the basis of this research.

This variability may be attributed to factors such as differences in sample characteristics and implementation conditions. It highlights the need to synthesize findings from various studies to derive a more generalizable reliability estimate for the Turkish adaptation of the BDI. This study aims to address this gap by providing a more precise and comprehensive understanding

of the scale's reliability. It is anticipated that the findings will not only contribute to a deeper understanding of the BDI's psychometric properties in Turkish contexts but also enable more accurate interpretation of studies utilizing this scale. A review of the literature reveals that while numerous meta-analyses focus on group compari-son types, such as transactional effectiveness and group differences (Gözübebek, 2012; Kablan *et al*., 2013; Şahin & Tekdal, 2005; Temel *et al*., 2020; Ulubey & Toraman, 2015), there is only one correlation meta-analysis study addressing the validity and reliability of diagnostic methods. This study, conducted by Eser and Aksu (2021), aimed to investigate the reliability generalization of the Beck Depression Inventory-II (BDI-II) adapted to Turkish culture.

Eser and Aksu included only English-language articles published between 2011 and 2019 in their meta-analysis. Using the Varying Coefficient Meta-Analytical Method (VC), they analyzed 40 studies and reported an average reliability level of .898 (Cronbach's alpha). However, their study had several limitations, including the exclusion of Turkish-language publications, a relatively small sample size, and a restricted time frame of 2011-2019. No meta-analyses have yet been conducted that address these limitations by examining a broader time range, incorporating a larger sample size, and including Turkish-language publications. This study aims to fill this gap in the literature, providing a more comprehensive and representative analysis of the reliability of the Beck Depression Inventory-II in Turkish contexts.

Based on this context, the primary aim of this study is to determine the reliability of the Beck Depression Scale, a widely used tool for diagnosing depression, a condition that significantly impacts an individual's life. The study employs meta-analysis, an advanced statistical technique, to achieve this objective. To address this overarching aim, the following research question was formulated: *"What is the reliability coefficient obtained by collectively analyzing the samples from studies that utilized the Beck Depression Scale?"*

2. METHOD

2.1. Research Model

This study is a quantitative research project designed using the meta-analysis method. Meta-analysis involves the synthesis and interpretation of the results from multiple studies on a specific subject by statistically combining their findings (Atalmış & Köse, 2018; Borenstein *et al*., 2009; Cooper *et al*., 2009). Field (2001) describes meta-analysis as a statistical technique based on the findings derived from the results of independent studies.

The following stages were followed in conducting the meta-analysis for this research (Borenstein *et al*., 2009; Hedges & Olkin, 1985; Higgins & Green, 2011; Higgins & Thompson, 2002; Sutton & Higgins, 2008):

1. Reviewing the literature
2. Determining the inclusion criteria
3. Coding
4. Calculating the effect size
5. Performing the heterogeneity test
6. Determining the model according to the heterogeneity test result
7. Calculating the Overall Effect Size
8. Interpreting the results

2.2. Data Collection Process

2.2.1. *Research sample*

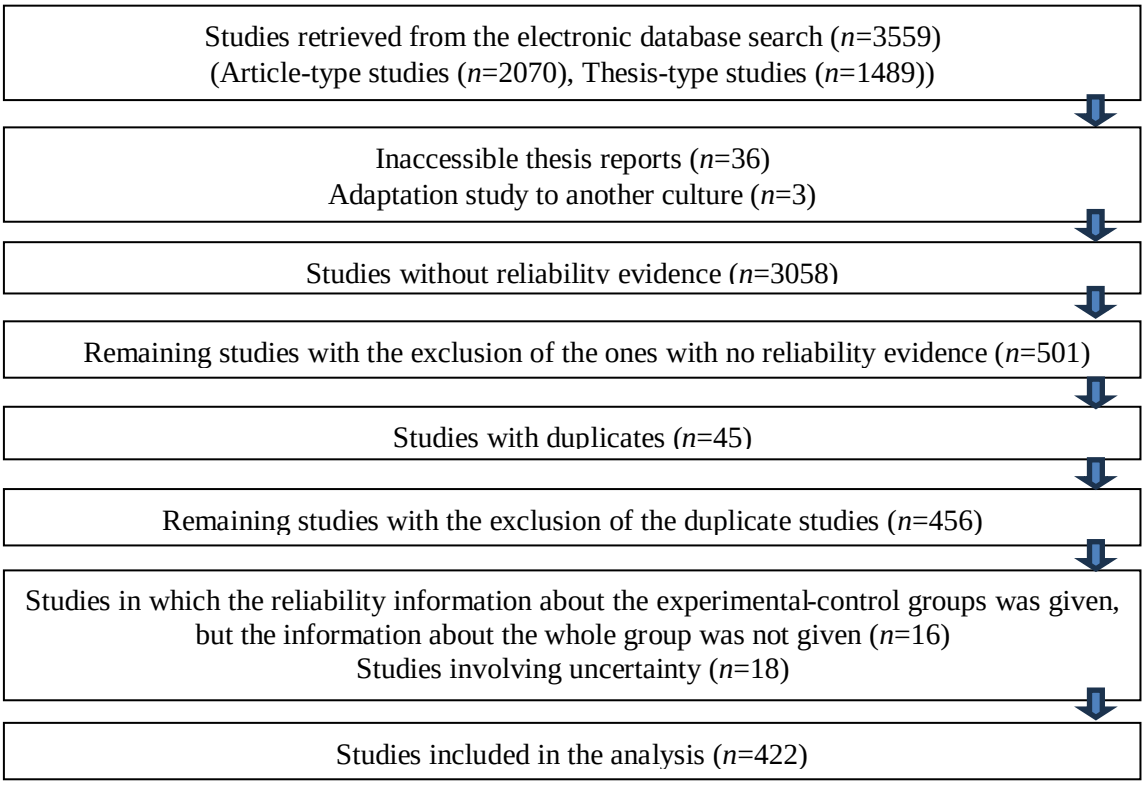
Studies utilizing the Beck Depression Scale were identified through a systematic search of the ProQuest (2021), HEC National Thesis Center (2021), and Google Scholar (2021) databases. The initial search was conducted on June 19, 2021, followed by a second search on July 31, 2021, to ensure completeness and accuracy. The keywords *"Beck Depression Inventory"* and *"Beck Depression Scale"* were used during the screening process.

In meta-analysis research, it is essential to establish clear inclusion criteria for the studies prior to conducting the analysis (Glass, 1976). For this study, the following criteria were applied to determine which studies would be included:

- Having been conducted in Turkey between 1961 and 2021
- Being a Master's / Doctoral thesis or an article published in peer-reviewed journals
- Having reported the reliability coefficient (Cronbach alpha) regarding the Beck Depression Scale
- Having open access permission of the full text of the studies

A total of 3,559 studies were retrieved from the electronic database search, with 41% sourced from the National Thesis Database and 59% from Google Scholar. Of these, 2,070 articles and 1,489 theses were assessed based on the predetermined criteria. Ultimately, 422 studies that reported the Cronbach's alpha coefficient as the internal consistency reliability estimation method were included in the analysis. The process of selecting these studies is depicted in the flowchart provided in Figure 1.

Figure 1. *Flowchart of data collection process.*



2.2.2. *Data coding*

To facilitate accurate data examination and minimize errors due to oversight, the researchers developed a coding form. This form included information such as the study's title, publication year, author details, and psychometric properties. To validate the coding form, consultations were held with two measurement experts. Based on their feedback, the form was refined, and its inter-coder reliability was evaluated using Kappa statistics.

The Kappa statistic assesses the degree to which the observed agreement between independent coders exceeds the agreement expected by chance (Sim & Wright, 2005). For this evaluation, 40 studies were randomly selected and independently coded by two researchers. The calculated Kappa value was .91, which, according to Landis and Koch (1977), indicates "perfect" agreement. These findings confirm that the coding form demonstrates high validity and reliability.

2.3. Data Analysis

Microsoft Excel was used for data preparation, while the Comprehensive Meta-Analysis (CMA) program was utilized for data analysis. The effect sizes for all studies included in the analysis were calculated, and the "average effect size" was derived from these calculations. Subsequently, a heterogeneity test (Q statistic) was conducted on the effect sizes to determine whether the Fixed Effects Model (FEM) or the Random Effects Model (REM) would be more appropriate for the analysis.

The Fixed Effects Model assumes that the effect size is equal and constant across all studies included in the analysis. In this model, variability arises solely from sampling error. Conversely, the Random Effects Model allows for the possibility that the effect sizes of individual studies may differ (Borenstein *et al.*, 2009). This model considers two sources of variability: sampling error at the individual level and random effects variance (Borenstein *et al.*, 2009; Cooper *et al.*, 2009; Şen & Yıldırım, 2020; Üstün & Eryılmaz, 2014). Card (2011) notes that results obtained using the Random Effects Model are generally more representative and widely applicable than those derived from the Fixed Effects Model.

In this study, bias analyses were performed following the selection of the appropriate model. The analyses included Fail-Safe N, Duval and Tweedie's Trim-and-Fill Method, and Egger's Regression Test. Funnel and forest plots were also examined to supplement and provide additional context to the test results. Effect sizes were interpreted using the criteria outlined by Thalheimer and Cook (2002), as shown in Table 1.

Table 1. *Interpretation of Cohen's d effect sizes.*

Cohen's *d* Benchmarks	Interpretation of Effect Size
[(-0.15) – 0.15)	Insignificant effect
[0.15 – 0.40)	Small effect
[0.40 – 0.75)	Moderate effect
[0.75 – 1.10)	Large effect
[1.10 – 1.45)	Extremely large effect
[1.45 -)	Strong effect

The statistic used to calculate effect size was Cohen's *d*, chosen for its computational simplicity, the availability of standardized reference ranges for interpretation, and the convenience these standards provide (Cohen, 1988). Additionally, Cohen's *d* was selected because the study involved a large sample size. Cohen's d is calculated by dividing the difference between group means by the pooled standard deviation.

3. FINDINGS

Meta-Analysis findings regarding the reliability level of Beck Depression Scale are presented below.

3.1. Heterogeneity Test

Before conducting the data analysis, the homogeneity or heterogeneity of the effect sizes from the included articles and theses was assessed using the Q-Test. The *Q* statistic is a widely used metric in meta-analysis to determine whether the variance between studies is significantly greater than what would be expected from random error alone. It evaluates the extent to which observed effect sizes deviate from the average effect size.

The *Q* statistic is interpreted alongside its degrees of freedom (*df*) and *p* value. If the *Q* value is high and the *p* value is low (typically $p < .05$), it indicates significant heterogeneity among the studies. However, it is important to note that a high *Q* value does not always confirm heterogeneity, as it can also result from small sample sizes. Therefore, it is recommended to

interpret the *Q* statistic in conjunction with other measures, such as the I^2 statistic, to provide a more comprehensive understanding of heterogeneity. The presence of heterogeneity influences the choice of the statistical model. If significant heterogeneity is detected, the Random Effects Model is preferred, as it accounts for variability both within and between studies. Conversely, if heterogeneity is low and all studies are assumed to represent the same effect size, the Fixed Effects Model may be appropriate (Borenstein *et al*., 2009; Higgins & Thompson, 2002).

In this study, the results of the Q-Test indicated significant heterogeneity among the effect sizes, leading to the selection of the Random Effects Model for the analysis. The detailed results of the heterogeneity test are presented in Table 2.

Table 2. *Findings regarding effect sizes (Pearson r) and heterogeneity test.*

Model	*k*	*SE*	*Z*	*p*	%95 CI		*Q*	*df*(*Q*)	*p*
					Lower Limit	Upper Limit			
Fixed	422	0.003	452.01	.00	1.37	1.38	4511.51	421	<.001
Random	422	0.010	134.12	.00	1.37	1.41			
Overall Effect Size	0.88								

Upon examining Table 2, the number of studies included in the analysis (*k*) is 422. The standard error (*SE*) of the effect size was calculated as 0.003 for the Fixed Effects Model and 0.010 for the Random Effects Model. The *Z* statistic was 452.01 for the Fixed Effects Model and 134.12 for the Random Effects Model. The heterogeneity test conducted to determine the appropriate model for calculating the overall effect yielded a *Q* value of 4511.51 ($p < .05$). Given this *Q* value, which was significant at 421 degrees of freedom (*df*) and a 95% confidence level, it was concluded that the studies exhibited significant heterogeneity. Consequently, the Random Effects Model was selected for the analysis. Using the Random Effects Model, the overall effect size was calculated as 0.88. According to Cohen's (1988) classification, this represents a large effect size. Therefore, the reliability coefficient for all studies included in the meta-analysis was determined to be 0.88, reflecting a high level of reliability for the Beck Depression Scale.

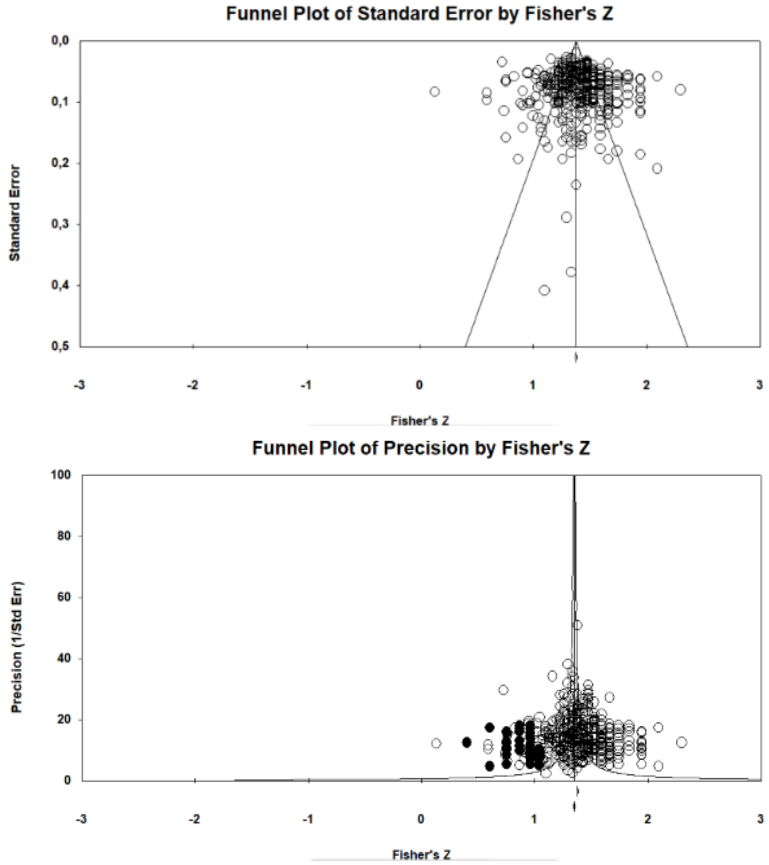
3.2. Publication Bias

Publication bias occurs when the studies included in a meta-analysis fail to represent the entirety of available studies in the literature (Rothstein *et al*., 2005). This bias is especially likely if the analysis exclusively includes studies with "statistically significant" findings while excluding those reporting nonsignificant results (Borenstein *et al*., 2009). To evaluate publication bias in this study, Funnel Scatterplots were analyzed alongside statistical tests such as Duval and Tweedie's Trim-and-Fill Test, Fail-Safe N, and Egger's Regression Test. Duval and Tweedie's Trim-and-Fill Test examines the symmetry of the study distribution in the meta-analysis. Asymmetric distributions may arise due to unpublished studies. This method corrects for bias by first "trimming" the studies that create asymmetry and then "filling" the gaps by imputing missing studies to achieve symmetry. The adjusted results provide a revised estimate of the effect size, accounting for the potential impact of unpublished studies (Duval & Tweedie, 2000).

Fail-Safe N assesses the robustness of meta-analysis findings by calculating the number of null-effect studies required to negate the significant results. Rosenthal's Fail-Safe N focuses on *p* values, determining the minimum number of additional studies needed to render the findings nonsignificant. Orwin's Fail-Safe N focuses on effect size, estimating the number of additional studies needed to reduce the observed effect size below a specified threshold. Higher Fail-Safe N values suggest that the meta-analysis results are more resistant to publication bias (Orwin, 1983; Rosenthal, 1979).

Egger's Regression Test uses a regression approach to evaluate the symmetry of effect sizes against standard errors. In a well-balanced meta-analysis, effect sizes are expected to form a symmetric funnel-shaped distribution. If the studies cluster asymmetrically in the Funnel Plot, it suggests the presence of publication bias (Egger *et al.*, 1997). The Funnel Scatterplots illustrating the distribution of effect sizes and indicating whether publication bias exists are presented in Figure 2.

Figure 2. *Funnel scatterplots for effect sizes.*



Upon examining Figure 2, the data in the Funnel Plot appear to exhibit a symmetrical distribution. According to Sterne *et al*. (2011), if the effect sizes, represented as dots for each study, are symmetrically distributed around the vertical line in the Funnel Plot, it suggests the absence of publication bias. In this case, the symmetrical distribution observed in Figure 2 supports the conclusion that publication bias is unlikely.

Furthermore, the Fail-Safe N value was calculated as 7,328, indicating that at least 7,328 null-effect studies would need to be added to invalidate the significant findings of this meta-analysis. This robust result provides additional evidence against the presence of publication bias. However, symmetrical distribution in the Funnel Plot and a high Fail-Safe N value alone may not provide conclusive evidence. To strengthen the findings, Duval and Tweedie's Trim-and-Fill Test results are presented in Table 3, offering further analysis to confirm the absence of publication bias.

Table 3. *Test results regarding publication bias.*

Variable	Duval and Tweedie's Trim-and-Fill Method		Egger's Regression Test (p)		
	Trimmed Study	Observed / Corrected	Deg. Free.	*t*	*p*
Studies	43	$1.39_{(1.36 - 1.41)}$ / $1.34_{(1.32 - 1.36)}$	42	1.44	.15

According to Duval and Tweedie's Trim-and-Fill Method, presented in Table 3, the number of studies that need to be trimmed to correct asymmetry in the Funnel Plot is determined, and the

effect size is recalculated after these studies are removed. If the difference between the observed and corrected effect sizes is not statistically significant, it provides evidence against the presence of publication bias (Pamuk *et al.*, 2015). In this study, the analysis revealed that 43 studies would need to be trimmed to achieve symmetry. However, the difference between the observed and corrected effect sizes was not statistically significant ($p = .15$; $p > 0.05$), indicating the absence of publication bias. Similarly, the results of Egger's Regression Test further support this conclusion, with a p value of .15 ($p > .05$). These findings collectively confirm that this study is free from publication bias, ensuring the robustness and reliability of its meta-analytic results.

3.3. Distribution of Confidence Intervals for Effect Sizes

The distribution of confidence intervals and study weights for the effect sizes of the research reports included in this analysis is presented in Table 4. Given the impracticality of listing all 422 studies in a single table, the results for a selection of studies are provided as representative samples.

Table 4. *Confidence intervals and study weights of effect sizes of studies included in the analysis.*

Study Name	Publication Date	Effect Size	Sd. Error	Var.	Confidence Intervals		*Z*	*p*
					Lower Limit	Upper Limit		
Ağca, Z.	2017	1.528	0.098	0.010	1.335	1.720	15.578	<.001
Ağdemir, B.	2017	1.293	0.071	0.005	1.155	1.432	18.336	<.001
Ahçıoğlu, A.	2020	1.157	0.061	0.004	1.038	1.276	19.079	<.001
Ak, G.	2001	1.333	0.089	0.008	1.159	1.507	15.023	<.001
Aka, B. T.	2011	1.333	0.044	0.002	1.248	1.418	30.603	<.001

The effect sizes and confidence intervals for each study are presented in Table 4, Figure 3, and Figure 4. The forest plots in Figure 3 display squares with horizontal lines extending on both sides, representing the confidence intervals. The horizontal position of each square corresponds to the effect size of the respective study. Larger squares indicate larger sample sizes, reflecting a greater contribution of that study to the overall average effect size. The horizontal lines represent the standard error, with longer lines indicating higher standard error values (Sen & Yıldırım, 2020). In Figure 3 and Figure 4, it is evident that most studies, except for those conducted by Yıldırım (2010), Yılmaz (2017) and Yılmaz (2019) exhibit relatively short lines, indicating low standard errors and high contributions to the overall analysis. Since study weight is closely related to sample size, it can also be inferred that the weights of these studies are similar.

Figure 3. *Forest plot regarding confidence intervals and overall effect.*

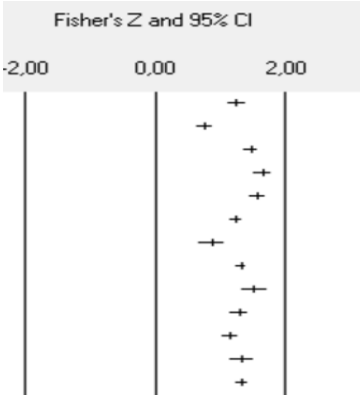
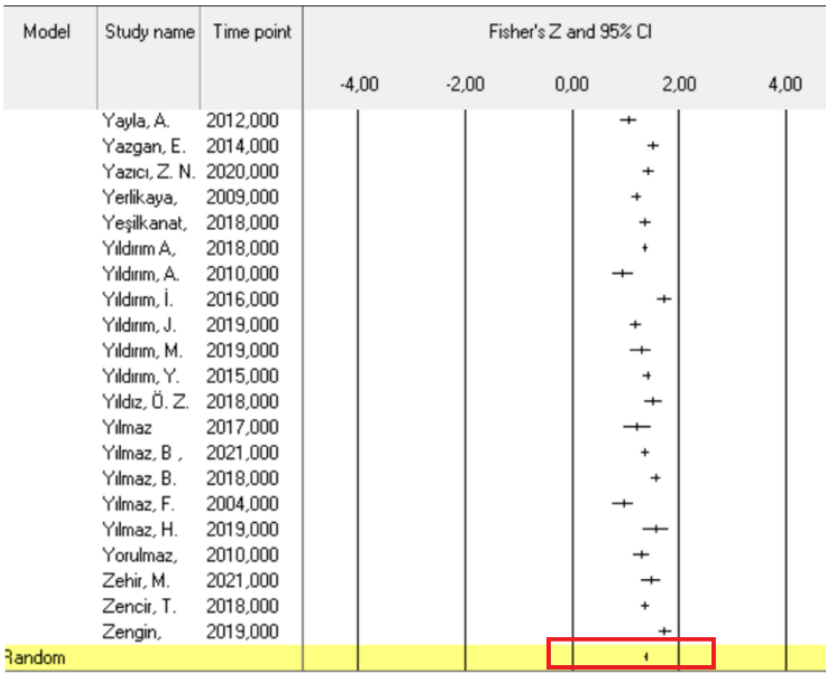


Figure 4. *Forest chart regarding studies.*



The "Random" line at the bottom of Figure 4 represents the overall meta-analysis result. Upon examining this line, it is evident that the confidence interval is quite narrow, suggesting high estimation precision of the effect sizes. Additionally, the average effect size is substantial, as indicated by the position of the line at the center.

4. DISCUSSION and CONCLUSION

This study aimed to derive a generalizable reliability coefficient for the Beck Depression Scale by synthesizing the reliability coefficients reported in primary studies conducted with the scale. Using meta-analysis, an advanced statistical method, the quantitative findings of similar studies were combined, enabling the generalization of the results to the population. Moreover, the meta-analysis approach mitigates the potential disadvantages arising from variations in sample sizes and implementation conditions across individual studies.

The analysis revealed that when the samples of the studies included in the meta-analysis were considered collectively, the reliability coefficient for the Beck Depression Scale was .88. This value demonstrates a very high level of reliability for the scale. These findings align with existing literature (Beck *et al.*, 1996; Steer & Clark, 1997). For instance, Dozois and Covin (2004) reported an average reliability coefficient of .91 in their analysis of 13 studies, while Eser and Aksu (2021), using the VC meta-analysis model with 40 articles, calculated a Cronbach's alpha value of .898. These results corroborate the findings of this study, further confirming the reliability of the Beck Depression Scale.

Regarding the time frame covered, Dozois and Covin (2004) analyzed studies published since 1996, while Eser and Aksu (2021) focused on studies published from 2011 onward. In contrast, this research spans a broader period, covering studies conducted between 1961 and 2021, thereby encompassing the time frames addressed in both of these prior studies. To assess potential bias in the study's results, Duval and Tweedie's Trim-and-Fill Test, Fail-Safe N, Egger's Regression Test, and funnel scatterplots were analyzed. These analyses concluded that no bias was present. With the exception of three studies - Yıldırım and Çevik (2010), Yılmaz (2017), and Yılmaz (2019) the weights of the studies included in the meta-analysis were found to be very similar. Additionally, the confidence interval observed in the forest plot for study weights was narrow, indicating high precision in the estimation of effect sizes.

The reliability coefficient calculated in this research was r = .88. This coefficient matched the reliability reported in 34 studies and was comparable to the findings of 149 studies that estimated reliability coefficients between .86 and .90 (Bozyel, 2017; Dağ, 2012; Kahveci, 2020; Keleş, 2019; Sevilmiş, 2019; Turan, 2020). However, this result differs from 11 studies that reported reliability coefficients below .70, such as those by Orhan (2019), Alan (2015), Türkmen (2019), and Abdulkerim (2019).

This discrepancy may not be attributable to the measurement tool itself but rather to differences in the characteristics of the samples or the conditions under which the scale was administered. These factors could significantly influence the reliability outcomes observed in different studies. This study focused on the Beck Depression Scale, a widely used tool for diagnosing depression. However, other psychological assessment tools were not considered. It is recommended that future research employ the meta-analysis method to assess and verify the reliability of other commonly used psychological diagnostic tools. While this study determined the reliability of the Beck Depression Scale using the meta-analysis method, it did not examine its validity. Future research could address this gap by conducting a meta-analysis that includes all available validity studies of the Beck Depression Scale. Additionally, this study employed the internal consistency reliability estimation method, using the Cronbach's alpha coefficient as a criterion for including studies. Future studies might consider evaluating the reliability of the Beck Depression Scale using other reliability estimation methods through meta-analysis.

It should also be noted that some studies remain unpublished because they yield nonsignificant results, a phenomenon referred to as the “file drawer problem” (Rosenthal, 1979). This issue raises the possibility that the studies included in this research may not represent all relevant studies in the literature. To address this, future research on the reliability of the Beck Depression Scale should strive to include and publish nonsignificant results to ensure a more comprehensive and unbiased representation of the available data.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Ankara University Social Sciences Ethics Committee, 27/06/2022-177.

Contribution of Authors

All authors have equally contributed to all section of this study. The authors read and approved the final manuscript.

Orcid

Yasemin Duygu Esen https://orcid.org/0000-0002-9300-0389
Filiz İsnaç https://orcid.org/0000-0002-2578-912X
H. Deniz Gülleroğlu https://orcid.org/0000-0001-6995-8223

REFERENCES

Abdulkerim, Ş. (2019). *Determination of depression and anxiety symptom levels and related factors in individuals with chronic obstructive pulmonary disease* [Unpublished master’s thesis]. Gazi University.

Abramson, J.H. (1994). *Making sense of data* (2nd ed.). Oxford University Pres.

Alan, H. (2015). *Effect of progressive muscle relaxation exercises and reflexology applied during chemotherapy treatment on pain, fatigue, anxiety, depression and gynecologic cancer patients' quality of life in randomized control study* [Unpublished master’s thesis]. Hacettepe University.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). American Psychiatric Association.

Arthur Jr, W., Bennett, W., & Huffcutt, A.I. (2001). *Conducting meta-analysis using SAS*. Psychology Press.

Atalmış, E.H., & Köse, A. (2018). Turkish prospective teachers' attitudes towards the teaching profession: a meta-analysis study. *Journal of Measurement and Evaluation in Education and Psychology*, *9*(4), 393-413. https://doi.org/10.21031/epod.410287

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.

Belsher, G., & Costello, C.G. (1988). Relapse after recovery from unipolar depression: a critical review. *Psychological Bulletin 104*, 84–96.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to meta-analysis*. Wiley.

Bozyel, G. (2017). *An examination of the mediating role of metacognition in the relationship between childhood trauma and psychological symptoms among undergraduate students* [Unpublished master's thesis]. Dokuz Eylül University.

Burcusa S.L., & Iacono W.G. (2007). Risk for recurrence in depression. *Clinical Psychology Review 27*, 959–985.

Card, N.A. (2011). *Applied meta-analysis for social science research*. The Guilford Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.

Cooper, H., Hedges, L.V., & Valentine, J.C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage.

Dağ, B. (2012). *Sleep behaviour of adolescents and the factors affecting it* [Unpublished master's thesis]. İstanbul University.

Demiray, P. (2013). *The effect of project based learning method: A meta analytic study* [Unpublished master's thesis]. Ankara University.

Dinçer, S. (2014). *Eğitim bilimlerinde uygulamalı meta analiz*. Pegem Akademi.

Dozois, D.J.A., & Covin, R. (2004). The Beck Depression Inventory-II (BDI-II), Beck Hopelessness Scale (BHS), and Beck Scale for Suicide Ideation (BSS). In M.J. Hilsenroth & D.L. Segal (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 2, pp. 50–69). John Wiley & Sons Inc.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89-98. https://doi.org/10.1080/01621459.2000.10473905

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. https://doi.org/10.1136/bmj.315.7109.629

Eser, M.T., & Aksu, G. (2021). Beck Depression Inventory-II: A study for meta-analytical reliability generalization. *Pegem Journal of Education and Instruction, 11*(3), 88-101. https://doi.org/10.14527/pegegog.2021.013

Field, A.P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed and random-effects methods. *Psychological Methods*, *6*(2), 161.

García-Batista, Z.E., Guerra-Peña, K., Cano-Vindel, A., Herrera-Martínez, S.X., & Medrano, L.A. (2018). Validity and reliability of the Beck Depression Inventory (BDI-II) in general and hospital population of Dominican Republic. *PloS one*, *13*(6). https://doi.org/10.1371/journal.pone.0199750

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8.

Gözüyeşil, E. (2012). *The effect of brain based learning on academic achievement: A meta analytic study* [Unpublished master's thesis]. Niğde University.

Hamidi, R., Fekrizadeh, Z., Azadbakht, M., Garmaroudi, G., Taheri Tanjani, P., Fathizadeh, S., & Ghisvandi, E. (2015). Validity and reliability Beck Depression Inventory-II among the

Iranian elderly population. *Journal of Sabzevar University of Medical Sciences, 22*(1), 189-198. https://doi.org/10.21859/jsums-2201189

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Academic Press.

Higgins, J.P.T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions.* Wiley-Blackwell. https://doi.org/10.1002/9780470712184

Higgins, J.P.T., & Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-1558.

Kablan, Z., Topan, B., & Erkan, B. (2013). The effectiveness level of material use in classroom instruction: A meta-analysis study. *Educational Sciences: Theory & Practice, 13*(3), 1629-1644. https://doi.org/10.12738/estp.2013.3.1692

Kahveci, M. (2020). *Sexual function status, depression and marital adjustment in women in the climacteric period* [Unpublished master's thesis]. İzmir Katip Çelebi University]

Keleş Gözütok, Ç. (2019). *The effect of anxıety, depression and fear of childbirth on labor results in last trımester pregnancy* [Unpublished master's thesis]. İstanbul Medipol University.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

McDermott, L.M., & Ebmeier, K.P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of Affective Disorders, 119*(1-3), 1-8. https://doi.org/10.1016/j.jad.2009.04.022

Monroe, S.M., & Harkness, K.L. (2012). Is depression a chronic mental illness? *Psychological Medicine, 42*(5), 899-902. https://doi.org/10.1017/S0033291711002066

Orhan, Ç. (2019). *Evaluation of depression and quality of life of women with breast cancer.* [Unpublished master's thesis]. Trakya University.

Orwin, R.G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*(2), 157-159.

Rosenthal, R. (1979). The 'file-drawer problem' and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641.

Rothstein, H.R., Sutton, A.J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* Wiley.

Steer, R.A., & Clark, D.A. (1997). Psychometric characteristics of the Beck Depression Inventory-II with college students. *Measurement and Evaluation in Counseling and Development, 30*(3), 128–136.

Şahin, M.C., & Tekdal, M. (2005). İnternet tabanlı uzaktan eğitimin etkililiği: Bir meta-analiz çalışması [The Effectiveness of Internet-Based Distance Education: A Meta-Analysis Study]. *Proceedings of the Academic Informatics 2005 Conference, 2*(4), 467–474.

Şen, S., & Yıldırım, İ. (2020). *CMA ile meta-analiz uygulamaları.* Anı Yayıncılık.

Sevilmiş, R. (2019). *The relationship between internet addiction and depression among university students* [Unpublished master's thesis]. Üsküdar University.

Sim, J., & Wright, C.C. (2005). The Kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257-268. https://doi.org/10.1093/ptj/85.3.257

Solomon, D.A., Keller, M.B., Leon, A.C., Mueller, T.I., Lavori, P.W., Shea, M.T., … Endicott, J. (2000). Multiple recurrences of major depressive disorder. *American Journal of Psychiatry, 157*, 229–233. https://doi.org/10.1176/appi.ajp.157.2.229

Sutton, A.J., & Higgins, J.P.T. (2008). Recent developments in meta-analysis. *Statistics in Medicine, 27*(20), 2077-2090. https://doi.org/10.1002/sim.3108

Temel, S., Yildiz, T., & Aslan, F.E. (2020). The effect of marital status on burnout levels of nurses: A meta-analysis study. *Journal of Clinical Medicine of Kazakhstan, 4*(58), 51-56. https://doi.org/10.23950/jcmk/vol04iss58/a9

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research articles: A simplified methodology. *Educational and Psychological Measurement, 62*(3), 503–514. https://doi.org/10.1177/0013164402042001

Turan, Y. (2020). *The effect of functional dependence of stroke patients on depression and quality of life* [Unpublished master's thesis]. Tekirdağ Namık Kemal University.

Türkmen, E. (2019). *Examination of the relationship between personality traits, anger expression styles and depressive symptoms in university* [Unpublished master's thesis]. Maltepe University.

Ulubey, Ö., & Toraman, Ç. (2015). The effect of creative drama on students' academic achievement: A meta-analysis study. *Mustafa Kemal University Journal of Social Sciences Institute, 12*(32), 195-220. https://doi.org/10.16992/mku.209027

Wang, Y.P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Brazilian Journal of Psychiatry, 35*, 416-431. https://doi.org/10.1590/1516-4446-2012-1048

Wiebe, J.S., & Penley, J.A. (2005). A psychometric comparison of the Beck Depression Inventory-II in English and Spanish. *Psychological assessment, 17*(4), 481. https://doi.org/10.1037/1040-3590.17.4.481

Yıldırım, A., & Çevik, B. (2010). Orem's self-care deficit theory-based assessment on the nursing interventions for a diabetic foot case. *Journal of Continuing Medical Education, 25*(5), 205-212. https://doi.org/10.23950/jcmk/vol25iss5/a9

Yılmaz, B.T. (2017). *Effects of aerobic exercises on the life quality, self-esteem and depression levels of the individuals with depression.* [Doctoral Dissertation, Cumhuriyet University].

Yılmaz, H. (2019). *Care burden, depression and related factors of caregivers with bipolar disorder* [Unpublished master's thesis]. Aydın Adnan Menderes University.

Zis, A.P., & Goodwin, F.K. (1979). Major affective disorder as a recurrent illness: A critical review. *Archives of General Psychiatry, 36*(3), 293-303.