International Journal of Progressive Education, Volume 16 Number 6, 2020
© 2020 INASED

A Review of Meta-Analysis Articles in Educational Sciences Conducted Between 2010 and 2019 in Turkey

Mehmet Taha Eser [i]
Akdeniz University

Meltem Yurtçu [ii]
Inonu University

Abstract

This study examines meta-analysis studies published in the field of educational sciences from 2010–2019, in journals indexed within the scope of the TÜBİTAK ULAKBİM TR Directory. Of the 163 studies scanned, 26 meta-analysis studies that meet the inclusion criteria constitute the sample of this content analysis study. Within this research, a meta-analysis control form (MACF) developed by the researchers was used. The coding made by each of the two researchers has been considered, and it is concluded that the coherence between the codings was is sufficient (82%). The results of this research indicate that the most obvious problems in meta-analysis studies are: not establishing hypotheses; not calculating reliability between encoders; not using a flow chart (in terms of traceability); the use of commercial software in the analysis; not combining effect sizes on a common metric; use of the fail-safe N in determining publication bias; evaluation of I^2 on the basis of categories; and decision making according to the result of the heterogeneity test in model determination. It is thought that the current study will contribute methodologically to the avoidance of errors in future meta-analysis studies.

Keywords: Meta-Analysis, Methodology, Content Analysis

DOI: 10.29329/ijpe.2020.280.2

----

[i] **Mehmet Taha Eser,** Instructor Dr., Assessment and Evaluation in Education, Educational Sciences, Akdeniz University, ORCID: 0000-0001-7031-1953

Correspondence: taha.eser@hacettepe.edu.tr

[ii] **Meltem Yurtçu,** Assist. Prof. Dr., Educational Sciences, Inonu University, ORCID: 0000-0003-3303-5093

INTRODUCTION

Many studies in the social sciences investigate the differentiation of the dependent variable according to the independent variable categories. The final result obtained from the inferences based on the significance of the frequencies varies, for several reasons (Aloe, Amo & Shanahan, 2014; Thiese, Ronna & Ott, 2016). Therefore, the same study may give different results, in different samples, under different conditions, or may be interpreted differently. This fact will lead to different tendencies among researchers in the evaluation of a subject in the literature. In such cases, the process of combining the studies together and reviewing them will give more meaningful results, in terms of guiding the literature. Most calculation procedures of the significance test require the outlined studies to be independent. Considering dependent studies as independent and interpreting the results accordingly may cause significance test errors (Rosenthal, 1995); this can be prevented by combining several non-independent studies as a single study through various dependent variables (Rosenthal, 1991). Meta-analysis provides statistical analysis of the results collected from more than one study (Aloe et al., 2014; Card, 2012); it also provides a method for arranging a large and inconsistent mass of findings and evaluating overall effect size (Cooper, Hedges & Valentine, 2009; Glass, 1976; Rosenthal, 1991; Rosenthal, 1995). Meta-analysis allows researchers to achieve more accurate and more reliable results than any primary study or non-quantitative narrative review can provide (Nakagawa, et al.,2017; Rosenthal & DiMatteo, 2001). DeCoster (2004) expressed the objective of meta-analysis as providing the same methodological discipline as the literature obtained from experimental studies. Başol, Doğuyurt and Demir (2016) stated that considering the technique, methodology and results together will lead to more subjective results. Therefore, meta-analysis can provide more accurate information to researchers, through an objective review of the studies carried out in the field of interest.

Meta-analysis is mostly used in scale development and descriptive studies (Başol et al., 2016); it is also preferred in studies evaluating the relationships between dependent and independent variables (Rosenthal & DiMatteo, 2001).

Meta-analysis studies are not only systematic literature reviews involving the process of blending the results, they are used to perform more specific and more sensitive reviews (Card, 2012; Cook et al., 1992; Hedges & Olkin, 1985; Higgins & Green, 2011).

Although the processes required for meta-analysis are basically the same, they are classified differently by different researchers. According to Mullen (1989), meta-analysis is performed in three steps: 1) central tendency, 2) heterogeneity of the results of the works, and 3) explanation of the diversity among them, with the properties of the works. DeCoster (2004), however, listed the steps as: 1) defining the relevant theoretical relationships, 2) collecting the set of works that provide data regarding the relationships, 3) coding the works and calculating the effect sizes, 4) analysing the effect of the moderator variables and the distribution of the effect sizes, and 5) reporting and interpreting the results.

The minimum number of works that can be included in a meta-analysis study is two, but the results from two works may show an unstable structure (Borenstein, Hedges, Higgins & Rothstein, 2009; Rosenthal, 1995). In meta-analysis studies, reviewing a small number of works poses a problem, and the quality of these works (whether good or bad) is criticised (Rosenthal & DiMatteo, 2001). Therefore, it is very important to confirm that the works to be included in the analysis address the same categories (Başol et al., 2016). This requirement, which aims to minimise deviation in the data set, depends on the comprehensiveness and systematicity of the inclusion and exclusion processes for the works to be covered within the scope of the study (Côté & Jennions, 2013; Higgins et al., 2019). For instance, using predefined questions and keywords and screening a single database with inclusion/exclusion criteria eliminate some biases and can be accepted as systematic techniques, although they are not comprehensive. For meta-analysis studies involving social sciences, it can be said that there is room for better documentation of search steps, in terms of systematicity and repeatability. The number of works to be included in/excluded from the meta-analysis is determined by the control forms, according to the specified criteria and methods in the literature (Başol et al., 2016; Rosenthal, 1991; Rosenthal & DiMatteo, 2001), increasing the internal reliability and external validity of the study (Başol et al., 2016). With a correct coding, therefore, it will be easier to explain changes in effect size (Furtak, Seidel, Iverson & Briggs, 2012; Lipsey & Wilson, 2001).

Effect size is the most important part of a meta-analysis reviews. If the sample is not too small, it provides illustrations of various central trend and variability indices, as well as effect sizes; this is often very valuable (Rosenthal, 1995; Rosenthal & DiMatteo, 2001). Two main effect size families are 'family *r*' and 'family *d*' (Rosenthal, 1995; Rosenthal & DiMatteo, 2001; Williams, Greenwood & Parker, 2013). The most important members of the first family are Pearson's r and Fisher's z transformation. In the ranking of the standard r for comparisons, meta-analysis generally uses the conversion of the correlation coefficients from Pearson's r to Fisher's z (Borenstein et al., 2009). The three most important members of the *d* family are Cohen's d, Hedges' g, and Glass's Δ (Rosenthal, 1995; Rosenthal & DiMatteo, 2001). Hedges' g is the standardised average difference that converts all effect sizes into a common metric that can be interpreted as the difference between the two groups, measured as a reference to a standard deviation (Cooper & Hedges, 1994). An advantage of Cohen's d and Hedges' g is that they are off-scale or standardized. The main limitation of both is their being parametric statistics (Williams et al., 2013).

Two basic effects are mainly used within the scope of meta-analysis: fixed and random effects. When the fixed effects model is adopted, the significance test is based on the total sample size, and the generalisation is carried out over the sample sizes. Hence, in the fixed effects model the statistical power is high, but the generalisation is limited (Borenstein, Hedges & Rothstein, 2007; Hill, Davis, Roos & French 2019), which may sometimes create limitations. Therefore, the random effects model is used. When the random effects model is adopted, the significance test is based on the total number of included works, not the total sample size, and generalisation goes beyond the specific works coming from the same population. Therefore, as the heterogeneity increases with the increase of generalisability, its statistical power decreases. Since the number of random samples in the works included in the meta-analysis is quite low, the random effects application should not target too much sensitivity (Rosenthal, 1995). The answer to the question of whether to use the fixed effects or random effects model in a meta-analysis relates to how sampling was performed. The sampling method used in the universe of active works (from which we collected the works) determines how we can generalise the results of the analysis. The fixed effects model assumes that the results of the works conducted in a population according to certain rules can be generalised to cases similar to this population and these rules. On the other hand, the random effects model assumes that we define a universe of works, and the actual effect size is different for each work in the universe. The works included in the analysis are assumed to be a random example of all works in the universe. Therefore, the random effects model can generalise the results of the analysis to all universes and cases included within the framework of the sampling (Borenstein, 2019; Borenstein et al., 2009).

Publication bias is one of the factors to be considered in meta-analysis that may lead to misinterpretation (Ioannidis & Trikalinos, 2007; Lin & Chu, 2018). Publication bias is a potential problem for any systematic review, because works reporting high effect sizes are more likely to be published than those reporting small effect sizes (Borenstein et al., 2009; Lipsey & Wilson, 2001; Rosenthal, 1991; Rothstein, Sutton & Borenstein, 2006; Vevea & Woods, 2005). Van Aert, Wicherts and van Assen (2019) grouped the methods used to review publication bias under two categories, namely used to test publication bias and the verification of the effect size for publication bias. They defined bias determination methods under these categories. The most commonly used bias determination methods are the fail-safe N, funnel plot, Egger's linear regression, test of excess significance, Begg and Mazumdar's rank correlation test, normal q–q graph and p-uniform's publication bias test (Başol et al., 2016; Borenstein et al., 2009; Darabi, Liang, Suryavanshi & Yurekli, 2013; Lipsey & Wilson, 2001; Vevea & Woods, 2005). The symmetry of the funnel plot (one of the most commonly used methods) indicates that there is no publication bias, whereas its skewness indicates the presence of bias (Darabi et al., 2013; Vevea & Woods, 2005). Başol et al. (2016) express not including all works in the literature in the study as a validity issue, whereas Aloe et al. (2014) state that the inclusion of published and unpublished works in a meta-analysis study does not seriously affect the publication bias, therefore the publication bias may not completely result from the sample.

© 2020 INASED

In meta-analysis, heterogeneity tests are measures of the differences between the works. Heterogeneity estimates the amount of variation among reviewed works, arising from the variables and from the external error (Higgins, Thompson, Deeks & Altman, 2003). These tests are reported using H, R, I^2, $\tau^2$and Q-values, which are considered to complement each other in the interpretation. I^2 is a simple percentage of the total variance; it should not be interpreted as the amount of heterogeneity (Borenstein, 2019). Unlike the Q-value, I^2 does not depend on the total number of works used to calculate the summary effect size. Therefore, I^2 is the preferred statistic when reporting the heterogeneity of a small group (Higgins et al., 2003). A high I^2 value indicates that variance is associated with factors other than the variable being investigated (Higgins et al., 2003). The Q statistic is a weighted variance that indicates whether all works come from a population with chi-square distribution. A significant Q statistic indicates that effect sizes come from different populations (heterogeneity). If the Q statistic is not significant, all works are assumed to have the same population effect, indicating the use of fixed effects models (Darabi et al., 2013). As the I^2 value, the Q statistic should never be used as the amount of heterogeneity (Borenstein, 2019). The H statistic emerges using Cochran's Q statistic. The parameter τ^2 is defined as the variance of the actual effect sizes.

The R^2 statistics are calculated together with τ^2 and the typical variance within the group (study) σ^2, similar to the H^2 statistics. H, R and I^2 statistics are not dependent on the number of studies. Therefore, it is recommended that one of these coefficients is presented with the tests performed (Higgins & Thompson, 2002).

The following are expressed as ways of addressing heterogeneity: conducting a citation review and meta-analysis; allowing or ignoring heterogeneity using fixed and random effects models while modelling the data; and conducting research to explain and eliminate heterogeneity (Higgins et al., 2003).

The properties recorded for each work and outlined in the study set can be used as moderator variables, that is, variables associated with the amount of the effect size achieved for different works (Rosenthal, 1995). The simplest way to analyse the moderators is to compare the average effect sizes in different subgroups that constitute moderator levels. The analysis of the moderators in meta-analyses allows further testing of the details of the theory and a better understanding of the research literature (Rosenthal & DiMatteo, 2001).

Reporting the prediction interval and confidence intervals along with the data of the works covered in the meta-analysis is an important issue. The prediction interval provides information on the range of the effect sizes on a single metric. In other words, it gives information about the range. The accuracy of the prediction interval is somewhat affected by the number of works. If the analysis includes at least ten works, the prediction interval is likely to be accurate enough to be useful. The confidence interval is a sensitivity index that shows how precisely we estimate the average effect size (Borenstein et al., 2009).

Purpose of the Study

Meta-analysis studies are expected to include many criteria, especially in terms of methodology. The purpose of this study is the evaluation of meta-analysis works conducted in educational sciences between 2010 and 2019 in Turkey, including the steps followed in terms of methodology and the comments made regarding the findings and results obtained. At the same time, this study aims to prevent both conceptual and statistical mistakes that may occur in future meta-analysis studies.

Research Questions

Within the scope of the study, meta-analysis works conducted in Turkey from 2010–2019 were evaluated, considering the following points, in terms of methodology:

© 2020 INASED

1. What are the criteria that are considered in the introduction parts of the studies?
2. What are the criteria that are considered in the methodology parts of the studies?
3. What are the criteria that are considered in the conclusion, discussion and suggestions parts of the studies?

METHODOLOGY

Research Model

The works included in the study were evaluated in terms of methodology, using content analysis. The main objective of content analysis is to reach the concepts and relationships that express the collected data in an understandable way (Yıldırım & Şimşek, 2006).

In the study, the introduction, method and conclusion parts of each meta-analysis work were reviewed and evaluated according to the specified criteria. In this regard, a descriptive process was followed. Descriptive studies are defined as involving a type of research that enables the revealing of the current situation as it is.

Sample

In the current study, meta-analysis papers published in the journals indexed within the scope of the TÜBİTAK ULAKBİM TR Index (ULAKBİM) were analysed, using the content analysis method. A total of 163 works were accessed in the field of educational sciences, as a result of a screening carried out in ULAKBİM with the keywords ‘Meta-Analysis’ and ‘Meta-Analiz’. Of those 163 works, 26 works that meet the inclusion criteria were evaluated within the scope of the study.

Data Collection Process

Regarding the identification of the works to be evaluated within the scope of the study, the first screening date was 7 January 2020, and the last screening date was 18 February 2020. Since there is not a separate ‘educational sciences’ database in the ULAKBİM system, the screening was carried out on the social sciences database. Although the current study is a content analysis study, the screening process is outlined by using the flowchart suggested for use in the PRISMA notification for systematic review and meta-analysis. The flowchart for the screening process is given in Figure 1.

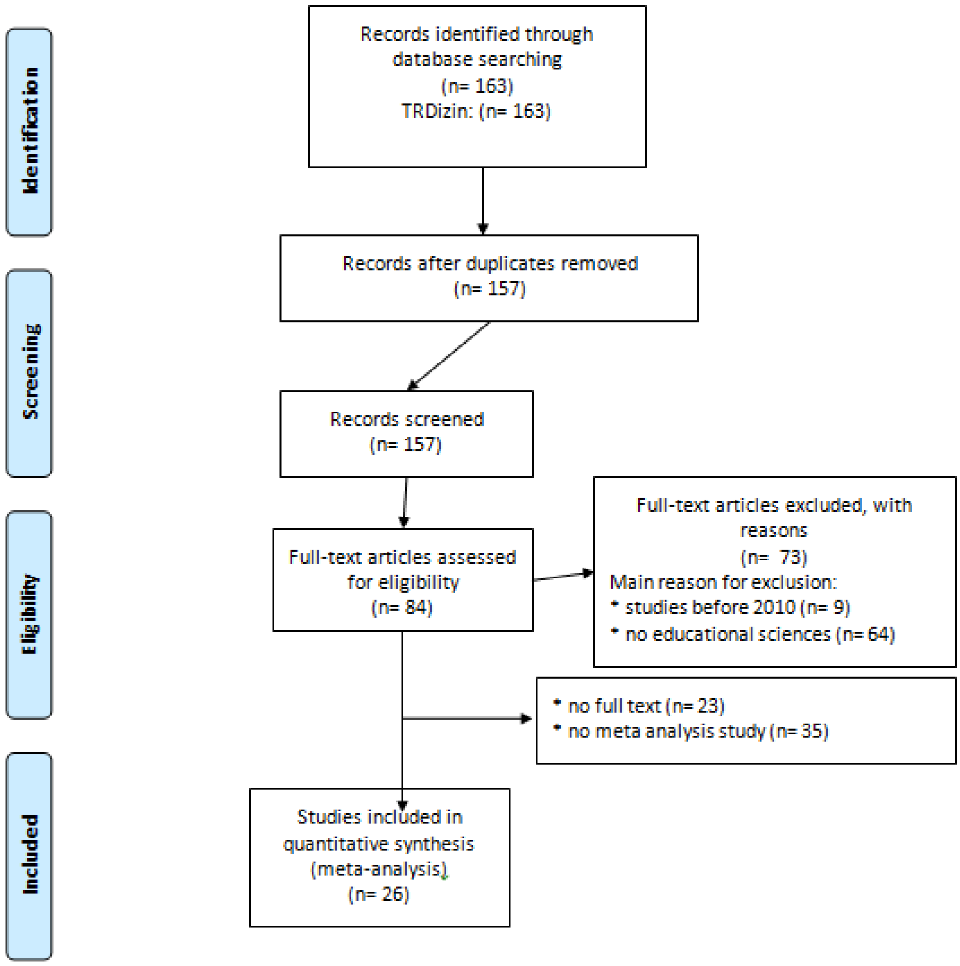

Figure 1

Data Collection Flowchart

In addition to the flowchart, Table 1 provides information about the journals in which the works evaluated in the study were published.

Table 1 Distribution of Papers Evaluated within the Scope of the Study, by Journal

Journal	f
Education and Science (Eğitim ve Bilim)	5
Turkish Studies: International Periodical for the Languages, Literature and History of Turkish or Turkic	4
Journal of Hacettepe University Education Faculty (Hacettepe Üniversitesi Eğitim Fakültesi Dergisi)	3
Journal of Abant İzzet Baysal University Education Faculty (Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi)	2
Eurasian Journal of Educational Research	1
Pegem Education and Training Journal (Pegem Eğitim ve Öğretim Dergisi)	1
Adıyaman University Journal of Educational Sciences (Adıyaman Üniversitesi Eğitim Bilimleri Dergisi)	1
Journal of İnönü University Education Faculty (İnönü Üniversitesi Eğitim Fakültesi Dergisi)	1
Kastamonu Education Journal (Kastamonu Eğitim Dergisi)	1
Journal of Erzincan University Education Faculty (Erzincan Üniversitesi Eğitim Fakültesi Dergisi)	1
SPORMETRE Journal of Physical Education and Sports Sciences (SPORMETRE Beden Eğitimi ve Spor Bilimleri Dergisi)	1
Journal of Trakya University Education Faculty (Trakya Üniversitesi Eğitim Fakültesi Dergisi)	1
Journal of Çukurova University Education Faculty (Çukurova Üniversitesi Eğitim Fakültesi Dergisi)	1
Journal of Mersin University Education Faculty (Mersin Üniversitesi Eğitim Fakültesi Dergisi)	1
Gazi University Journal of Gazi Education (Gazi Üniversitesi Gazi Eğitim Dergisi)	1
Educational Management in Theory and Practice (Kuram ve Uygulamada Eğitim Yönetimi)	1

© 2020 INASED

Regarding Table 1, five of the works were published in *Education and Science (Eğitim ve Bilim),* four works in *Turkish Studies: International Periodical for the Languages, Literature and History of Turkish or Turkic,* three works in the *Journal of Hacettepe University Education Faculty (Hacettepe Üniversitesi Eğitim Fakültesi Dergisi)* and two in the *Journal of Abant İzzet Baysal University Education Faculty (Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi).*

Regarding the evaluation of other features related to the papers but not included in Table 1: 4 of the 26 works were observed to focus on the effectiveness of the relationships and 22 on the effectiveness of the differences between the studies reviewed. Six works were from 2015, seven from 2016, seven from 2017, four from 2018 and two from 2019.

Validity and Reliability of the Study

Transactional content analysis for detailing screening forms is one of the methods that increase the validity of content analysis studies (Başol et al., 2016). In order to ensure high research validity, the items specifying the properties that meta-analysis should have (in terms of methodology) are included in the coding form. Each work was enumerated by the researchers to ensure that the works were included in the sample only once.

In order to determine the reliability of the coding processes to be carried out by the researchers, 6 of the 26 papers were selected at random and analysed by both researchers. The reliability of the coding was determined, and the agreement between the raters (expected to be at least 80%) was calculated as 82%, using the formulas suggested by Miles and Huberman (1994). In addition, to increase the reliability of the paper classification process, the classifications were discussed by the researchers and disagreements were resolved.

Data Collection Tool

Prior to the content analysis, the researchers developed a meta-analysis control form (MACF) to reveal the methodological deficiencies of the included meta-analysis works. PRISMA (2009), control forms developed by Geyskens, Krishnan, Steenkamp and Cunha (2009) and the meta-analysis evaluation form developed by Başol et al. (2016) were used in the development of the MACF. The characteristics of the meta-analysis process and the most common problems in meta-analysis were considered in its creation.

The items included in the MACF require that the encoder codes the questions related to the methods and techniques used in the study as Yes/No, except for one item. For some items, the researchers were expected to make explanations about the relevant case, in addition to the coding. The MACF consists of three sections: Introduction Part of the Study (14 items), Methodology Part of the Study (12 items), and Conclusion, Discussion and Suggestions Parts of the Study (3 items).

Data Analysis

The data obtained by the MACF were explained using frequency and percentage values.

Limitations

The study is limited to 26 studies available in the ULAKBİM system. The inclusion criterion for the works evaluated within the scope of the research was set as being a meta-analysis study. Following screening performed with the keywords 'Meta-Analysis' and 'Meta-Analiz', works lacking any statistical information among the meta-analysis were excluded from this research, according to the MACF.

© 2020 INASED

FINDINGS

Findings and Interpretations Regarding the Introduction Parts of the Studies

The findings obtained as a result of the analysis of 26 papers according to the first 14 items of the MACF form are given below.

Table 2 Descriptive Statistics of the Answers given to the Items in the Introduction Part

MACF items	f	%
The purpose of the study is explicitly stated.	25	96
The reason(s) for the selection of variables is explicitly stated.	24	92.3
The research problem is reported.	20	77
Sub-problems related to the research problem are reported.	16	62
One or more research hypotheses are established.	1	4
The databases in which the screening was performed are specified.	25	96
The dates of the screening process are reported.	12	47
The keywords used in screening are reported.	23	88
Inclusion criteria are reported.	26	100
Exclusion criteria are reported.	6	23
Unpublished studies are evaluated within the scope of the work.	17	65
Intercoder reliability is tested.	14	54
The number of screened works, eligible works and the ones included in the review are reported using a flowchart as well as the works exclude from the study and exclusion reasons.	3	12
The studies included in the meta-analysis are reported in a table.	10	38

Table 2 shows the satisfaction of the criteria in the MACF developed by the researchers (mainly based on the PRISMA criteria) by the 26 works published in 2010–2019. Works carry different properties that meet different criteria. For the reviewed papers, the criteria most frequently met are: reporting of inclusion criteria, reporting of keywords used in screening, specification of databases where screening was performed, explicit statement of the purpose of the study, explicit statement of the reason(s) for variable selection and reporting the research problem.

While 16 works contain information on both the start and end dates of the screening process, in 4 works (~ 15%) only the final screening date is included.

In addition, there are 12 works in which the studies to be included in the meta-analysis were not analysed by at least two coders. Therefore, for these, inter-coder reliability was not tested. Of the 14 works in which inter-coder reliability was tested, 9 (~ 64%) used the Miles–Huberman method, while 4 (~ 29%) employed Cohen's Kappa. In 1 (~ 4%) work, only the establishment of agreement is mentioned; no numerical value is given.

The criteria least satisfied by the 26 works are: establishment of a hypothesis; reporting of exclusion criteria; and reporting numbers of screened works, eligible works, works included in the review and works excluded from the study (with exclusion reasons) (i.e. having a flowchart). Among these, lack of exclusion criteria and absence of a flow chart affect the transparency of the research (Moher, Liberati, Tetzlaff & Altman, 2009).

Findings and Comments Regarding the Methodology Parts of the Studies

The findings obtained as a result of the analysis of 26 papers according to 12 items included in the MACF form are given below.

Table 3 Descriptive Statistics of the Answers given to the Items in the Methodology Part

MACF items	f	%
Software used in the analysis:	CMA (22) MetaWin (5) Excel (2) SPSS (3)	85 19 8 12
Normal distribution was tested to check whether combining the effect sizes of the studies is appropriate or not.	4	15
Information about the analysis of extreme values is provided.	0	0
Effect sizes were combined on a single metric using the conversion formula.	1	8
Effect sizes were interpreted and reported as small/medium/large.	26	100
Fail-safe N is reported.	19	73
Publication bias is specified.	25	96
Heterogeneity/homogeneity is specified.	26	100
Moderator analyses were performed.	20	77
The result of the heterogeneity test was taken into consideration in the selection of the model used.	2	8
Confidence intervals were interpreted.	7	27
Prediction intervals were interpreted.	0	0

As can be seen in Table 3, the CMA program was most frequently used in the works. In four of the works (~ 15%), no information is given about the statistical program used.

The effect sizes were observed to be interpreted as low/medium/high in all works. In 10 of the 26 works (~ 38%), Cohen's d was used in effect size calculations, whereas in 8 (~ 31%), Hedge's g was used. In eight works (~ 31%), the coefficient used in effect size calculations is not reported. The reasons for using the selected coefficient are reported only in one work.

Regarding the detection of publication bias, the funnel plot was used in 22 works (~ 85%); Egger's test in 5 (~ 19%), Begg and Mazumdar's rank correlation test in 3 (~ 12%) and Duval and Tweedie's method in 4 (~ 15%). No results related to the test of publication bias is reported in any of the works. In addition, in 11 of the 26 works (42%), fail-safe N is not reported, whereas in 11 (73%) of the remaining 15 works, fail-safe N was used to specify publication bias. It should be noted that five different coefficients were used for the specification of publication bias in two different works.

In 18 of the 26 works (~ 69%), both I^2 value and the Q test (heterogeneity or homogeneity test) were employed; whereas only the Q test was used in 8 works. In addition, in 16 works where I^2 was used, this coefficient was evaluated in three categories, as low, medium or high (25%, 50% and 75%, respectively).

At the model selection point, the random effects model according to heterogeneity was used in 20 of the 26 works (~ 77%), both fixed and random effects models were used in 1 work, the random effects model according to the objectives in the literature was used in 2 works and the random effects model according to both literature and heterogeneity was used in 1 work. In one work, the random effects model was reported to be preferred because of failing to meet common effect size assumption. In another work, the random effects model was reported to be preferred because the studies included in the meta-analysis were not functionally equal, and the work aimed to make a generalisation from the determined effect size. As can be seen from these results, it can be said that within the scope of the analysed works, model selection was made according to the results of the heterogeneity test.

In 20 of the 26 works (~ 77%), the structure formed by the effect sizes was found to be heterogeneous, and moderator analyses were performed. In two works, moderator analyses were reported to be performed according to the homogeneity of the structure formed by the effect sizes of the works included in the meta-analysis. In one work, the structure formed by the studies included in the meta-analysis was determined to be heterogeneous and moderator analyses were carried out, but no further information was provided in terms of revealing the reasons for heterogeneity. In another work, the structure formed by the effect sizes was found to be heterogeneous, but the moderator analysis was not performed; similarly, in another, the structure formed by the effect sizes was found to be homogeneous and the moderator analysis was not performed. In yet another work, the structure formed by the effect sizes was found to be heterogeneous, the moderator analysis was not performed and the absence of a moderator analysis was reported as a limitation of the study and it was suggested that it was performed.

Regarding the methodology parts of the analysed works, the most common shortcomings are as follows: prediction intervals were not interpreted, normal distribution was not tested to assess the appropriateness of combining effect sizes, the information regarding the analysis of extreme values was not provided and effect sizes were not combined on a single metric using the conversion formula. These can be interpreted as showing that the characteristics of the studies reviewed in the works were not substantially taken into consideration.

Findings and Comments Regarding the Conclusion, Discussion and Suggestions Parts of the Studies

The findings obtained as a result of the analysis of Conclusion, Discussion and Suggestions parts of 26 papers according to three items of the MACF are given below.

Table 4 Descriptive Statistics of the Answers Given to the Items in the Conclusion, Discussion and Suggestions Parts

MACF items	f	%
Limitations of the work are reported.	11	42
Generalisations are made considering the findings.	26	100
Adequate suggestions are submitted for future research.	14	54

As seen in Table 4, limitations are reported in 11 of the 26 works (~ 42%). The lack of reporting of the limitations in 15 of the works (~ 58%) can be interpreted as a negative factor. In all 26 works, generalisations are made considering the findings. In 11 of the 26 studies (~ 42%), suggestions for future research are submitted.

DISCUSSION, CONCLUSION AND SUGGESTIONS

Regarding the publication years of the works analysed in the study, no published study satisfying inclusion criteria was found between 2010 and 2014. This shows that interest in and the popularity of meta-analysis in the field of educational sciences have increased in the last five years. The vast majority of the works included in the study focus on the effectiveness of the differences, while the remaining few focused on the effectiveness of the relationships.

Research hypotheses were observed not to be established in the majority of the works. Considering that hypotheses explain the available data, suggest solutions to the problem and are testable (Karasar, 2011), the importance of establishing a hypothesis is clear.

Inter-coder reliability was found to be untested in a significant portion of the works reviewed in the study. Miles and Huberman (1994) emphasise that the performance of coding by two researchers on the same data set makes descriptions clearer; in addition, it is possible to reach an agreement about the meaning of the coding and the parts of the data belonging to a specific code if two researchers code the same data set.

Most of the works covered in this study contain deficiencies in the reporting of exclusion criteria, numbers of screened works, eligible works, works included in the review and excluded works (and the reasons for exclusion) by using a flow diagram, and in the reporting of the works included in the meta-analysis explicitly as a table. Additionally, the screening dates are not reported in some of the works. The lack of reporting of relevant details creates problems in terms of the transparency, reliability and repeatability of the research. At the same time, these deficiencies cause problems related to validity.

Regarding the computer programs used, the CMA program was observed to be used in most of the 26 works. CMA is a commercial software for meta-analysis applications. Meta-analyses were observed to be carried out by a single program in the reviewed works, which restrains the revealing of other evidences related to the results by using different outputs that may be obtained through other software.

Another remarkable deficiency is that normal distribution and extreme value assumptions were not tested in a significant number of the 26 works. The test of the assumptions before performing statistical analysis helps to ensure the accuracy of the analysis results and is one of the factors that eliminates the risk of the study deviating from scientific principles. Without this, the study becomes a non-scientific study.

In most of the works, the effect sizes were not combined on a common metric using transformation formula (or no information is given on this issue). The main objectives of the meta-analysis can be stated as: combining relevant studies and calculating the effect sizes, combining these calculated effect sizes on a common metric and, finally, achieving an average effect size from the effect sizes combined on a common metric (Cohen & Manion, 2001; Field, 2001).

Regarding the works reviewed in the course of the research, the coefficient used in effect size calculations is not mentioned in some works, and only one work provides information about why the coefficient used in the calculation of the effect size was preferred. Regarding the literature, the use of Hedge's g is recommended for very small sample sizes (< 20); for sample sizes greater than 20, it is suggested that the results will be approximately equal regardless of whether Cohen's d or Hedge's g is used (Durlak, 2009; Ellis, 2010; Hedges & Olkin, 1985). Failure to provide information about the type of the effect size and the reasons for preferring it is considered a threat to the differentiation of study results in meta-analysis.

Regarding the fail-safe N, no quantitative output is reported in some of the 26 studies reviewed in this research; the fail-safe N was used to address publication bias in some others. The fail-safe N is one of the expressions of the p value obtained from meta-analysis. It describes the robustness of a significant result by calculating how many studies with effect size zero can be added to the meta-analysis before the result loses statistical significance. The fail-safe N value is strongly affected by the average mediating effect of unpublished studies (Iyengar & Greenhouse, 1988). The fail-safe N is usually very large, even if the null hypothesis (H_0) is accepted. If the p value is too small, the fail-safe N will be too large; if the p value is slightly less than 0.05, the fail-safe N will be small (Hagger, Wood, Stiff & Chatzisarantis, 2010; Rosenthal, 1979). It gives information about the robustness of a statistically meaningful result, that is, the power of meta-analysis. Using a fail-safe N in determining the presence of publication bias may lead to false results (Higgins & Green, 2011; Scargle, 2000).

In the vast majority of works, the I^2 value was classified as low, medium or high in terms of heterogeneity, which is not a correct approach, as I^2 is not an absolute distribution index, but a ratio; it therefore does not give information about the distribution, and a heterogeneity that can be considered high in one context can be considered low in another (Borenstein, 2019; Harrer, Cuijpers, Furukawa & Ebert, 2019). Also, some works were observed to use only the Q test regarding heterogeneity. Since Q and I^2 are two different statistics with different superiorities, sticking to a single statistical result while making inferences about heterogeneity may lead to misinterpretations (Borenstein, 2019; Card, 2012).

Regarding heterogeneity, it is worth mentioning that heterogeneity cannot be interpreted as 'good' or 'bad' in its natural form. If our objective is to report that intervention raises scores to a certain value, we can say that, in this case, heterogeneity is indeed a problem. In the absence of heterogeneity, a common effect that applies to all populations can be reported. In the presence of heterogeneity, there is no common effect size and hence no common effect can be reported; however, in such a case, the degree of heterogeneity can be assessed, and effect sizes of 0.05 for some populations and as high as 0.95 for other populations can be reported. What makes heterogeneity 'good' or 'bad' varies according to the objective of the study (Borenstein, 2019; Higgins et al., 2003).

In the majority of the works reviewed in this study, the model selection was made according to the results of the heterogeneity test, which is not a correct approach, for several reasons. It is not logical to use the result of a statistically insignificant heterogeneity test in deciding whether the studies reviewed within the scope of a meta-analysis share a common effect. Although heterogeneity tests are considered to be important by the researcher, they may not be statistically significant. Also, the fixed effects model is used when such an approach is adopted. If a default model must be chosen, it would be more logical to choose the random effects model, which requires fewer assumptions about the consistency of the effects (Borenstein, 2019; Borenstein et al., 2009; Huedo-Medina, Sanchez-Meca, Marín-Martínez & Botella, 2006).

Confidence intervals were not interpreted in the majority of the works reviewed in this research. The confidence interval related to the variance between studies is of great importance for the random effects model, in terms of measuring the uncertainty of point estimates (Jackson & Bowden, 2016). Confidence intervals and point estimates are interpreted differently by the fixed effects and random effects models (Higgins & Green, 2011). Regarding the review of the works in terms of interpretation of prediction intervals, no data related to the relevant information was found. Prediction intervals show the actual effect size range of future studies. Failure to interpret confidence intervals and prediction intervals may increase the risk of making mistakes in the conclusions made for relevant studies and future studies and may damage the validity of the cases concerned.

In addition, study limitations are not reported in more than half of the reviewed works. Not reporting the limitations of a meta-analysis harms the repeatability of the study in terms of research strategy (Moher et al., 2015).

In view of the findings of this study, the following suggestions are issued for future studies:

Given that various types of meta-analysis exist (e.g. multilevel meta-analysis, Bayesian meta-analysis, network meta-analysis based on graph theory and structural equation modelling meta-analysis), it is recommended that different types of meta-analysis studies involving these are conducted.

It is recommended that meta-analyses are performed using open-source software such as R and free and open-source software such as Jamovi or JASP, where the analysis is carried out through drop-down menus.

In order to prevent the loss of internal reliability of meta-analysis, it is strongly recommended that inter-coder reliability is checked. Also, if there are likely to be many guesses, the use of a statistic like kappa is recommended for inter-coder reliability; however, if the coders are well-trained and less likely to guess, the use of statistics such as agreement rate is recommended (McHugh, 2012).

Regarding information about the presence or absence of publication bias, the use of p-uniform, Egger's test or Begg's test is recommended (rather than fail-safe N). Additionally, employing more than one method will strengthen the study, in terms of the reviewability of the results.

Regarding the selection of a fixed effects or random effects model, the decision should be taken with consideration of whether the studies examined within the scope of the research share a common effect size, as well as the objective of the meta-analysis. Model selection is related to the distribution of effect sizes and the consideration of related error sources. If the reviewed studies comprise only published works, the random effects model may be suggested as a more reasonable option (Borenstein et al., 2009).

Regarding the reporting of systematic review and meta-analysis studies in the literature, the use of the PRISMA flow chart, a review template in which decisions for the inclusion and exclusion of the relevant works are shown in a transparent way, is recommended (Moher et al., 2009; Moher et al., 2015)

In future meta-analysis studies, it is recommended that prediction intervals are reported, to show the real effect range of those studies.

REFERENCES

Aloe, A. M., Amo, C. L., & Shanahan, M. E. (2014). Classroom management self-efficacy and burnout: a multivariate meta-analysis. *Educ Psychology Review. 26*(1)*,* 101-126. doi:10.1007/s10648-013-9244-0

Borenstein, M. (2019). *Common mistakes in meta-analysis and how to avoid them.* Englewood, NJ: Biostat.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex, England: Wiley.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). *Meta-analysis fixed effect vs. random effects*. Retrieved from: https://www.meta-analysis.com/downloads/M-a_f_e_v_r_e_sv.pdf

Başol, G., Doğuyurt, M. F., & Demir, S. (2016). Türkiye örnekleminde meta analiz çalışmalarının içerik analizi ve metodolojik değerlendirilmesi. *International Journal of Human Sciences*, *13*(1), 714–745. doi:10.14687/ijhs.v13i1.3460

Card, N. A. (2012). *Applied meta-analysis for social science research.* New York, NY: Guilford Press.

Cohen, L., & Manion, L (2001). *Research Methods in Education* (5th ed.). New York: Routledge Falmer.

Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. A., & Mosteller, F. (1992). *Meta-analysis for explanation: A casebook.* New York, NY: Russell Sage Foundation.

Cooper, H., & Hedges, L. (1994). *The handbook of research synthesis*. New York, NY: Russell Sage.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis.* New York: Russell Sage Foundation.

Côté, I. M., & Jennions, M. D. (2013). The procedure of meta-analysis in a nutshell. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *The handbook of meta-analysis in ecology and evolution* (pp. 14–24). Princeton: Princeton University Press.

Darabi, A., Liang, X., Suryavanshi, R., & Yurekli, H. (2013). Effectiveness of online discussion strategies: A meta-analysis. *American Journal of Distance Education*, *27*(4), 228–241. doi:10.1080/08923647.2013.837651

DeCoster, J. (2004). *Meta-analysis notes*. Retrieved February, 2, 2020 from http://www.stat-help.com/meta.pdf

Durlak, J. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*(9), 917–928.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results.* Cambridge: Cambridge University Press.

Eser, M. T., Yurtçu, M., & Aksu, G. (2020). R Programlama Dili ve Jamovi ile Meta-Analiz Uygulamaları. Ankara: Pegema Yayıncılık.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, *6*(2), 161–180, doi: 10.1037/1082-989x.6.2.161

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research, 82*(3), 300–329. doi: 10.3102/0034654312457206

Geyskens, I., Krishnan, R., Steenkamp, J.-B. E. M., & Cunha, P. V. (2009). A review and evaluation of meta-analysis practices in management research. *Journal of Management, 35*(2), 393–419. https://doi.org/10.1177/0149206308328501

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

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136*(4), 495–525. https://doi.org/10.1037/a0019486

Harrer, M., Cuijpers, P., Furukawa, T. A, & Ebert, D. D. (2019). *Doing meta-analysis in R*: *A hands-on guide.* Retrieved from https: //bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/

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

Higgins J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions* (version 5.1.0). Retrieved from www.handbook.cochrane.org

Higgins J., Thomas J., Chandler J., Cumpston M., Li T., Page M.J., & Welch V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). Chichester, England: John Wiley & Sons.

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

Higgins, J., Thompson, S., Deeks, J., & Altman, D. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 2003*(327)*,* 557–560.

Hill, D. T., Davis, A. P., Roos, J. M., & French, M. T. (2019). Limitations of fixed-effects models for panel data. *Sociological Perspectives.* https://doi.org/10.1177/0731121419863785

Huedo-Medina, T., Sanchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 Index? *Psychological Methods, 11*(2), 193–206. doi:10.1037/1082-989X.11.2.193.

Ioannidis, J., & Trikalinos, T. (2007). The appropriateness of asymmetry tests for publication bias in meta-analysis: A large survey. *Canadian Medial Association Journal, 176*(8), 1091–96.

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science*, *3*(1), 109–117. Retrieved from www.jstor.org/stable/2245925

Jackson, D., & Bowden, J. (2016). Confidence intervals for the between-study variance in random-effects meta-analysis using generalised heterogeneity statistics: Should we use unequal tails? *BMC Medical Research Methodology, 16*(1), 118. doi: 10.1186/s12874-016-0219-y

Karasar, N. (2011). *Bilimsel araştırma yöntemleri*. Ankara: Nobel Yayınları.

Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics, 74*(3), 785–794. doi:10.1111/biom.12817.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Miles, M, B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. doi: 10.1371/journal.pmed.1000097

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4, 1. https://doi.org/10.1186/2046-4053-4-1

Mullen, B. (1989). *Advanced BASIC meta-analysis*. Hillsdale, NJ: Erlbaum.

Nakagawa, S., Noble, D. W. A., Senior, A. M., & Lagisz, M. (2017). Meta-evaluation of meta-analysis: Ten appraisal questions for biologists. *BMC Biology*. 15, 18. https://doi.org/10.1186/s12915-017-0357-7

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage. https://doi.org/10.4135/9781412984997

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*(2), 183–192. https://doi.org/10.1037/0033-2909.118.2.183

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology.* 52, 59–82. doi: 10.1146/annurev.psych.52.1.59

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* West Sussex, England: Wiley.

Scargle, J. D. (2000). Publication bias: The 'file drawer' problem in scientific inference. *Journal of Scientific Exploration*, *14*(1), 91–106.

Thiese, M. S., Ronna, B., & Ott, U. (2016). P value interpretations and considerations. *Journal of Thoracic Disease, 8*(9), 928–931.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10*(4), 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2019). Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PLoS ONE 14*(4): e0215052. https://doi.org/10.1371/journal.pone.0215052

Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*(1), 39–55.

Yıldırım, A., & Şimşek, H. (2006). *Sosyal bilimlerde nitel araştırma yöntemleri* (6. baskı). Ankara: Seçkin Yayıncılık.

REFERENCES EXAMINED WITHIN THE SCOPE OF THE RESEARCH

Ayaz, M. F. (2015). Probleme dayalı öğrenme yaklaşımının Öğrencilerin fen derslerindeki akademik Başarılarına etkisi: bir meta-analiz çalışması. *Turkish Studies International Periodical for the Languages, Literature and History of Turkish or Turkic, 10*(3), 139–160.

Ayaz, M. F., & Söylemez, M. (2016). The effect of project-based learning approach on students' attitudes towards science lesson: a meta-analysis study. *Adıyaman University Journal of Educational Sciences, 6*(1), 112–137.

Ayaz, M. F., & Söylemez, M. (2016). Proje tabanlı öğrenme yaklaşımının Türkiye'deki öğrencilerin fen derslerindeki akademik başarılarına etkisi: bir meta-analiz çalışması. *Eğitim ve Bilim, 178(*40), 255–283.

Aydın, A., & Sarıer, Y. (2013). Eğitim kurumu müdürlerinin liderliği ile okul çıktıları arasındaki ilişkilerin meta-analiz yöntemleri ile incelenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 13*(2). 257-275.

Aytaç, T. (2015). Öğretmenlerin okulda karşılaştıkları mobbinge ilişkin hizmet süresinin etkisi: Bir meta-analiz. *Kuram ve Uygulamada Eğitim Yönetimi, 21*(2), 161–182. doi:10.14527/kuey.2015.007

Batdı, V. (2015). A meta-analytic study regarding the effect of MIT on achievement, attitude and retention. *Turkish Studies International Periodical for the Languages, Literature and History of Turkish or Turkic,10*(7)*,* 197–218. doi:10.7827/TurkishStudies.8267.

Batdı, V. (2017). Eğitlence uygulamalarının akademik başarıya etkisi: meta-analitik bir çalışma. *GEFAD/GUJGEF, 37*(1), 47–62.

Baysal, Y. E., Mutlu, F., & Kış, A. (2019). Çağdaş yaklaşımlara dayalı laboratuvar uygulamalarının Türkiye'deki öğrencilerin laboratuvar derslerindeki tutumlarına etkisi: Bir meta-analiz çalışması. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 19*(3), 802–817.

.Çırak-Kurt, S., Yıldırım, İ., & Cücük, E. (2018). Harmanlanmış öğrenmenin akademik başarı üzerine etkisi: Bir meta-analiz çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 33*(3): 776–802. doi:10.16986/HUJE.2017034685

Çikrikçi, Ö. & Erzen, E. (2016). Cinsiyetin Akademik Erteleme Üzerindeki Etkisi: Meta-Analiz. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, *31(4):* 750-761, doi: 10.16986/HUJE.2016015387

Dağyar, M., & Demirel, M. (2015). Probleme dayalı öğrenmenin akademik başarıya etkisi: Bir meta-analiz çalışması. *Eğitim ve Bilim, 181*(40), 139–174.

Ergen, B., & Kanadlı, S. (2017). The effect of self-regulated learning strategies on academic achievement: A meta-analysis study. *Eurasian Journal of Educational Research*, *69*, 55–74. http://dx.doi.org/10.14689/ejer.2017.69.4

Haylı, R., G., Durmuş, E., & Kış, A. (2017). Evlilik doyumunun evlenme biçimi açısından İncelenmesi: Bir meta-analiz çalışması. *International Periodical for the Languages, Literature and History of Turkish or Turkic 12*(28), 429–444.

Kaldırım, A., & Tavşanlı, Ö. F. (2018). İş birlikli öğrenme yaklaşımının türkiye'deki öğrencilerin Türkçe derslerindeki akademik başarılarına etkisi: bir meta-analiz çalışması. *Eğitim ve Bilim*, *194*(43), 185–205.

Kansızoğlu, H. B., & Bayrak-Cömert, Ö. (2017). The effect of the process approach on students' writing success: A meta-analysis. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi,* 46(2), 541-586, doi:10.14812/cuefd.292046

Karadağ, E., İşçi, S., Öztekin, S., & Anar, S. (2016). Okul iklimi ve akademik başarı ilişkisi: Bir meta-analiz çalışması. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, *17*(2), 107–122. doi:10.17679/iuefd.17294446

Kasapoğlu, F., & Kış, A. (2016). Öznel iyi oluşun cinsiyet açısından incelenmesi: Bir meta-analiz çalışması. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 12*(3), 770–782.

Saraç, H. (2018). Yapılandırmacı yaklaşım öğrenme halkası modellerinin öğrenilen bilgilerin kalıcılığına etkisi: Meta analiz çalışması. *Kastamonu Education Journal, 26*(3), 753–764. doi:10.24106/kefdergi.413322

Sarı, İ. (2019). *Antrenörden algılanan özerklik desteği ve sporcuların özerk güdülenmesi arasındaki ilişki: Bir meta-analiz çalışması. 17(2),* 110-125. doi:10.33689/spormetre.455760

Satıcı, S. A., & Can, G. (2017). Cinsiyetin sosyal bağlılık üzerindeki etkisi: Bir meta-analiz çalışması. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi, 19*(2), 1-18. doi:10.17556/erziefd.289023

Sidekli, S., & Çetin, E. (2017). Okuduğunu anlama stratejilerinin okuduğunu anlamaya etkisi: Bir meta-analiz çalışması. *Trakya University Journal of Education Faculty*, *8*(2). 285-303, doi:10.24315/trkefd.321757

Şad, S. N., Kış, A., Demir, M. & Özer, N. (2016). Meta-analysis of the relationship between mathematics anxiety and mathematics achievement. P*egem Eğitim ve Öğretim Dergisi, 6*(3), 371-392

Şad, S. N., Kış, A.,& Demir, M.(2017). A Meta-Analysis of the Effect of Contemporary Learning Approaches on Students' Mathematics Achievement. Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education) 32(1): 209-227

Ural, G., & Bümen, N. (2016). Türkiye'de fen ve teknoloji öğretiminde yapılandırmacılığın öğretimsel uygulamaları üzerine bir meta-analiz. *Eğitim ve Bilim, 41*(185), 51–82.

Yılmaz, K., Altınkurt, Y., & Yıldırım, H. (2015). Türkiye'deki öğretmenlerin örgütsel vatandaşlık davranışlarına cinsiyet, kıdem ve branş değişkenlerinin etkisi: Bir meta analiz çalışması, *Eğitim ve Bilim, 178*(40), 285–304.

© 2020 INASED

Appendix A: Meta-analysis control form

Introduction Part of the Study

1. The purpose of the study was explicitly stated
   (0) No (1) Yes
2. The reason (s) for the selection of variables were explicitly stated
   (0) No (1) Yes
3. Research problem was reported
   (0) No (1) Yes
4. Sub-problems related to the research problem were reported
   (0) No (1) Yes
5. Research hypothesis (s) were established
   (0) No (1) Yes
6. The databases in which the screening was performed were specified
   (0) No (1) Yes
7. The dates of the screening process were reported
   (0) No (1) Yes
8. The keywords used in screening were reported
   (0) No (1) Yes
9. Inclusion criteria were reported
   (0) No (1) Yes
10. Exclusion criteria were reported
   (0) No (1) Yes
11. Unpublished studies were evaluated within the scope of the work
   (0) No (1) Yes
12. Intercoder reliability was tested
   (0) No (1) Yes
13. The number of screened works, eligible works and the ones included in the review were reported using a flowchart as well as the works excluded from the study and exclusion reasons.
   (0) No (1) Yes
14. The studies included in the meta-analysis were reported in a table
   (0) No (1) Yes
15. Softwares used in the analysis
   …………………
16. Normal distribution was tested to check whether combining the effect sizes of the studies is appropriate or not.
   (0) No (1) Yes
17. Information about the analysis of extreme values was provided
   (0) No (1) Yes
18. Effect sizes were combined on a single metric using the conversion formula
   (0) No (1) Yes
19. Effect sizes were interpreted and reported as small-medium-large
   (0) No (1) Yes
20. Fail-safe N was reported
   (0) No (1) Yes
21. Publication bias was specified
   (0) No (1) Yes
22. Heterogeneity/Homogeneity was specified
   (0) No (1) Yes
23. Moderator analyzes were performed
   (0) No (1) Yes
24. The result of the heterogeneity test was taken into consideration in the selection of the model used
   (0) No (1) Yes
25. Confidence intervals were interpreted
   (0) No (1) Yes
26. Prediction intervals were interpreted
   (0) No (1) Yes
27. Limitations of the work were reported
   (0) No (1) Yes
28. Generalizations were made considering the findings
   (0) No (1) Yes
29. Sufficient suggestions were submitted for future research
   (0) No (1) Yes