ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2023; 14(3); 249-265

A New Weighting Method in Meta-Analysis: The Weighting with Reliability Coefficient*

Yıldız YILDIRIM**   Şeref TAN***

Abstract

This study aimed to investigate the impact of various weighting methods for effect sizes on the outcomes of meta-analyses. For this purpose, a representative meta-analysis example examining the effect of the 5E teaching method on academic achievement in science education was discussed. Two effect size weighting methods were explored: one based on the inverse of the sampling error variance and the other utilizing the reliability of measures in primary studies. The study also assessed the influence of including gray literature on the meta-analysis results, considering factors such as high heterogeneity and publication bias. The research followed a basic research design and drew data from 112 studies, encompassing a total of 149 effect sizes. An exhaustive search of databases and archives, including Google Scholar, Dergipark, HEI Thesis Center, Proquest, Science Direct, ERIC, Taylor & Francis, EBSCOhost, Web of Science, and five journals was conducted to gather these studies. Analyses were performed by utilizing the CMA v2 software and employing the random effects model. The findings demonstrated divergent outcomes between the two weighting methods—weighting by reliability coefficient yielded higher overall effect sizes and standard errors compared to weighting by inverse variance. Ultimately, the inclusion of gray literature did not significantly impact any of the weighting methods employed.

Keywords: weighting methods, meta-analysis, reliability coefficient, gray literature

Introduction

Today, with the development of technology and the increase in globalization, science has become more rapidly developing and shared than in the past. As it is known, one of the essential features of scientific research is that it is reproducible and progresses cumulatively. The literature shows that many studies have been conducted in different fields within the framework of the same or similar research problems. For this reason, while there was no need to combine the findings in the past because the number of studies was less, over time, it has become necessary to combine these studies in many fields because of the increase in the number of studies conducted within the same framework and the repetition of studies. As a result, this necessity led to the birth of the meta-analysis method.

The method used to combine findings from repeated studies has a long history (Hedges & Olkin, 1985). Simpson and Pearson's (1904) study was one of the first examples of meta-analysis and evaluated the effectiveness of smallpox vaccine (National Research Council, 1992). Since studies are frequently repeated, it has led to the development of statistical techniques for combining results in different fields. The combining estimates from different studies were not used much in educational or psychological research until Glass proposed it in 1976 because, in studies conducted in these fields, certain psychological constructs or variables were not measured on the same scale in all studies. In 1976, Glass suggested using the effect size index to combine the results of studies conducted with different scales, making the studies comparable and combinable regardless of which scale was used (Hedges & Olkin, 1985). Glass (1976), the eponymist (Mutluer et al., 2020), called the combination of research findings in his study meta-analysis.

* This study is a part of doctoral thesis conducted under the supervision of Prof. Dr. Şeref TAN and prepared by Yıldız YILDIRIM GÖRGÜLÜ

** Assist. Prof. Dr., Aydin Adnan Menderes University, Faculty of Education, Aydin-Türkiye, e-mail: yildizyldrm@gmail.com, ORCID ID: 0000-0001-8434-5062
*** Prof. Dr., Retired from Gazi University, Faculty of Education, Ankara-Türkiye, sereftan4@yahoo.com, ORCID ID: 0000-0002-9892-3369

To cite this article:
Yıldırım, Y., & Tan, Ş. (2023). A new weighting method in meta-analysis: the weighting with reliability coefficient. *Journal of Measurement and Evaluation in Education and Psychology, 14*(3), 249-265. https://doi.org/10.21031/epod.1351485
Received: 28.08.2023
Accepted: 26.09.2023

In meta-analysis, an overall effect size is calculated by non-weighting or weighting the effect sizes of primary studies (Fuller & Hester, 1999). To calculate the overall effect size, summing the effect sizes of the primary studies and dividing by the total number of studies, i.e., averaging the effect sizes, is a method used mainly in the past and is called non-weighting in the literature. In addition to the average effect size (overall effect size) without weighting, there are different weighting methods in the literature. These methods generally assume that the error arises from the sample and are based on sample size and sampling error variance. Weighting the effect sizes in primary studies by sample size to obtain the overall effect size was proposed by Hunter and Schmidt (1990). Hunter and Schmidt (2004) stated that if the effect size in the population was assumed to be fixed across studies, to make the best estimation of this effect size, it is necessary to work not with the arithmetic mean of the studies but with a weighted average in which each effect size was weighted by the sample size in the study. Hedges and Vevea (1998) proposed a method called inverse-variance weighting, in which the effect sizes of primary studies are weighted by the inverse of the sampling error variance. In this method, the calculation of weights varies according to random effects and fixed effects models. In the random effects model, in addition to the sampling error variance, the between-studies variance is also taken into account. There are studies on the effects of weighting methods in the literature (Englund et al., 1999; Marín-Martínez & Sánchez-Meca, 2009; Schmidt et al., 2009; Shuster, 2010; Yıldırım & Şahin, 2023). In these studies, the effects of methods such as non-weighting, weighting by sample size, and weighting by the inverse of the sampling error variance were compared and examined.

In meta-analysis studies in the literature, primary studies are generally weighted by the inverse of the sampling error variance based on the sample size, and it is assumed that the error variance is caused only by the sample. However, there are sources of error variance other than the sample. The reliability coefficient is an index that also includes other sources of random error. The error can be caused by the measurement tool or the individual performing the measurement, as well as the environment in which the measurement is made and the construct of the trait. Rosenthal (1991) also stated that it is wise to weight studies in proportion to the quality of the studies using any weight between zero and one.

Based on the research on weighting in the literature, this study, unlike other studies, aimed to examine how the overall effect size and standard error obtained from the meta-analysis were affected by weighting with the reliability coefficient in addition to weighting with the inverse of the sampling error variance because assuming that the error is caused only by the sample is not exactly the right approach. No other study using weighting with a reliability coefficient was found in the literature. Using the reliability coefficient in synthesizing studies in meta-analysis and weighting effect sizes is this study's original and innovative aspect that will contribute to the literature. In this respect, the study differs from other methodological meta-analysis studies. The study discusses how these weighting methods change the results of meta-analysis. The research is essential since not many studies in the literature use a different weighting technique other than weighting by sampling error variance. In addition, the fact that weighting by reliability is used for the first time in this research by formulating weighting by reliability coefficient makes the research essential.

In the literature, it is frequently observed that meta-analysts in educational research do not include unpublished studies such as papers, reports, and theses (Altunoğlu et al., 2020; Bozdemir et al., 2017; Yeşilpınar Uyar & Doğanay, 2018). Such studies are called gray literature. In addition to this situation, it has been observed that there are also studies that include only theses in meta-analysis studies (Alacapınar & Ok, 2020; Basit, 2020; Başpınar, 2021; Saraç, 2018). However, there are meta-analyses that included both published and unpublished studies (e.g., Fabiano et al., 2021; Toraman et al., 2018; Özdemir, 2023). For this reason, it is another question of how the inclusion and exclusion of gray literature in meta-analysis studies affect the meta-analysis results. Based on this, how the inclusion of gray literature under different weighting methods affects the meta-analysis results is also examined within the scope of this study. Although there are studies in the literature that examine the effect of the inclusion of gray literature (Hartling et al., 2017; Moher et al., 1996), what makes this research different from other studies is that it examines this effect in the context of two weighting methods. This study is essential since reviewing the impact of gray literature under different weighting methods is a new issue.

Aim

This study aims to examine how the meta-analysis results are affected when the studies are weighted by sampling error variance and reliability in examining the effect of the 5E teaching method on academic achievement in science education by meta-analysis. In addition, within the scope of the research, it is also examined how the inclusion and exclusion of gray literature affect the meta-analysis results when weighting is done by sampling error variance and reliability in examining the effect of the 5E teaching method in science education on academic achievement by meta-analysis.

Method

Research Model

In this study, meta-analysis was conducted by using the weighting method with the reliability coefficient, which is different from the weighting method with the inverse of the sampling error variance since the error in measurement and evaluation processes is not only caused by the sample. Thus, a new weighting method was proposed to find a solution to the existing problem. According to Karasar (2013), basic research aims to add new knowledge to existing knowledge, and there are different levels of basic research. These are explication, elaboration, determination of cause-effect relationship, and theory development levels. A study at the explication level tries to determine exactly what an existing problem is, what variables are affected by it, and what the most appropriate approaches to explain the situation might be. In this context, the research is at the explication level of the basic research type. On the other hand, it was also examined how the inclusion of gray literature in meta-analysis studies affected the results of meta-analysis when the methods of the inverse of sampling error variance and weighting with reliability were considered. From this point of view, the research also has a descriptive purpose since an existing situation is tried to be revealed.

Data Collection Process

Primary studies constitute the study data in meta-analysis. In the meta-analysis study to be conducted, the study data consists of the studies to be selected according to the determined criteria. In order to strengthen this meta-analysis study methodologically, PICO (Participant/Population, Interventions, Comparisons, Outcomes) was followed. According to PICO, we need to determine which participants, interventions, control groups/comparisons, and outcomes will be taken into account and which we are interested in when constructing the problem. (Higgins & Green, 2008). Therefore, databases were searched with the keywords given in Table 1 to select primary studies to be included in the meta-analysis. In addition, the journals in Table 1 were also included in the search.

Table 1
Databases, keywords and number of studies

Databases	Keywords	Number of Studies
Google Scholar	“5E” + “fen” + “başarı”	1678
Dergipark	5E AND fen AND başarı	61
HEI Thesis Center	5E AND fen AND başarı	125
Proquest	5E AND fen AND başarı	37
Science Direct	5E AND fen AND başarı	0
Science Direct	5E AND science AND achievement	84
ERIC	5E AND fen AND başarı	0
ERIC	5E AND science AND achievement	47
Taylor & Francis	5E AND fen AND başarı	0

Taylor & Francis	5E AND science AND achievement	261
EBSCOhost	5E AND fen AND başarı	268
EBSCOhost	5E AND science AND achievement	130
Web of Science	5E AND fen AND başarı	0
Web of Science	5E AND science AND achievement AND Turkey	53
Journal Name		
Science Education		930
Journal of Research in Science Teaching		1036
Journal of Science Teacher Education		696
International Journal of Science and Mathematics Education		1149
Studies in Science Education		157
Total		6712

The databases presented in Table 1 were selected because these databases are frequently used in meta-analysis studies in the field of education (Arık & Yılmaz, 2020; Batdı & Batdı, 2015; Becker & Park, 2011; Lazonder & Harmsen, 2016; Sosa et al., 2011; Warfa, 2016 and Xie et al., 2018). The journals in Table 1 were selected because they have a high impact factor in the field. The databases were searched with relevant keywords, and all articles in the journals were searched without using keywords, and their full texts were analyzed. These full texts were analyzed according to the criteria determined. The criteria for selecting the study data for the meta-analysis study are listed as follows:

i. The period should be between January 2005 - December 2020,

ii. Papers, articles, dissertations, reports, etc., must have been conducted in a sample of Turkey,

iii. Designed as a weak experimental design, quasi-experimental design, true experimental design, or one of the mixed methods research that used one of the experimental designs in the quantitative research step,

iv. The language of publication must be Turkish or English,

v. Primary studies must have been conducted at the 4th, 5th, 6th7th, 8th, 9th, 10th,11th, or 12th grade or at a higher education level and must be in the field of science, physics, chemistry and biology,

vi. The teaching in the treatment group must have been done with the 5E teaching model or with the 5E teaching model supported by additional applications,

vii. In the control group, traditional methods such as lecture, question and answer, discussion, demonstration, exhibition etc., must have been used, and if not stated in the study, when the authors were contacted via e-mail/message, it was confirmed in their response that they used traditional methods.

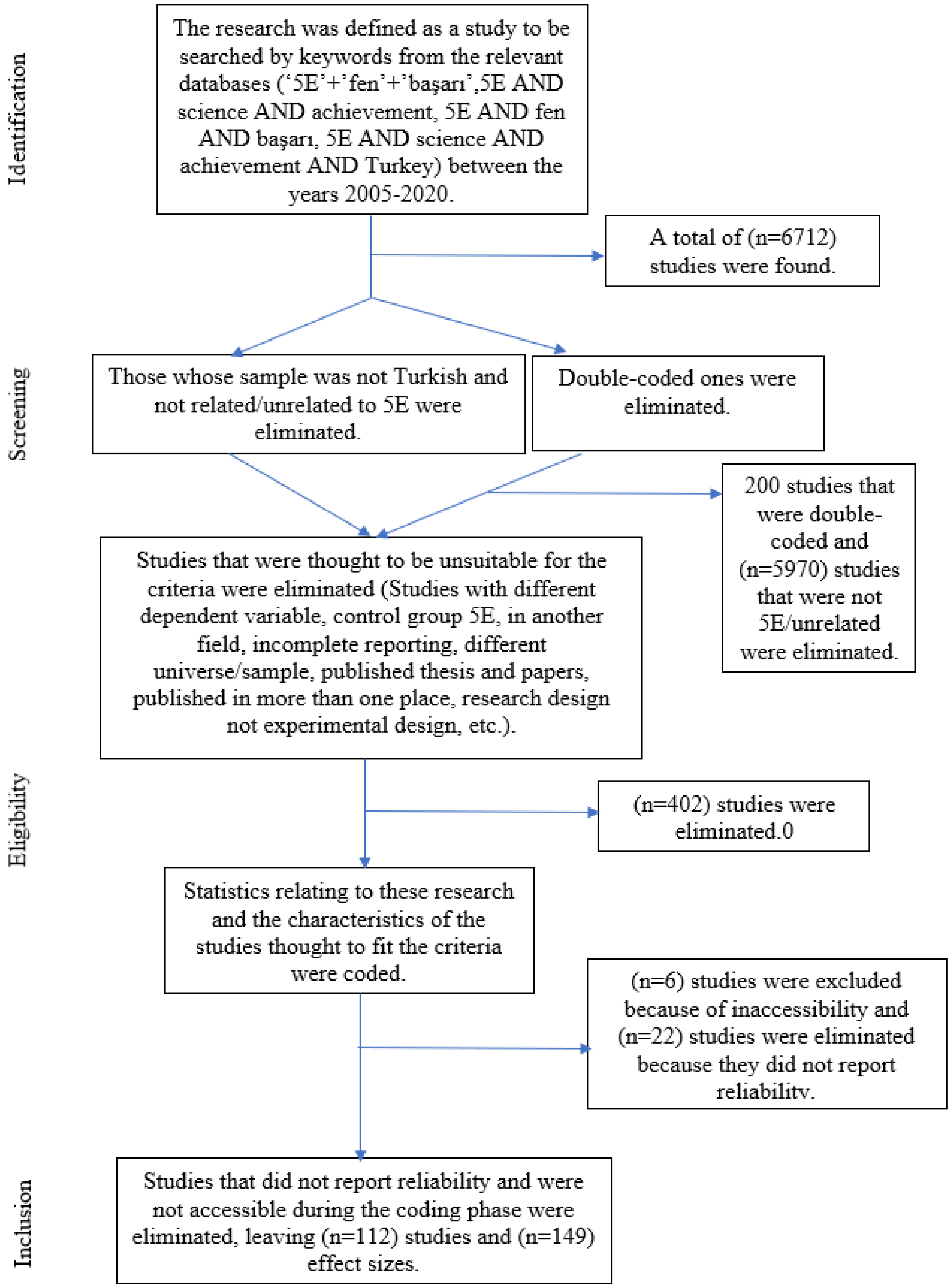
viii. As a data collection tool, tests such as multiple-choice achievement tests, concept tests, conceptual understanding tests, tests composed of open-ended items, and concept maps, which measure academic achievement and report reliability scores, must have been used.

ix. The dependent variable must be academic achievement or concept knowledge.

x. Report sufficient quantitative data and sample size to allow calculation of the effect size.

Primary studies to be included in the meta-analysis were identified according to the search criteria made with keywords in the databases. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for the process of identifying these studies is given in Figure 1 (Liberati et al., 2009).

Figure 1
PRISMA flowchart


Identification
Screening
Eligibility
Inclusion
The research was defined as a study to be searched by keywords from the relevant databases ('5E'+'fen'+'başarı',5E AND science AND achievement, 5E AND fen AND başarı, 5E AND science AND achievement AND Turkey) between the years 2005-2020.
A total of (n=6712) studies were found.
Those whose sample was not Turkish and not related/unrelated to 5E were eliminated.
Double-coded ones were eliminated.
200 studies that were double-coded and (n=5970) studies that were not 5E/unrelated were eliminated.
Studies that were thought to be unsuitable for the criteria were eliminated (Studies with different dependent variable, control group 5E, in another field, incomplete reporting, different universe/sample, published thesis and papers, published in more than one place, research design not experimental design, etc.).
(n=402) studies were eliminated.0
Statistics relating to these research and the characteristics of the studies thought to fit the criteria were coded.
(n=6) studies were excluded because of inaccessibility and (n=22) studies were eliminated because they did not report reliabilitv.
Studies that did not report reliability and were not accessible during the coding phase were eliminated, leaving (n=112) studies and (n=149) effect sizes.

According to the PRISMA flowchart in Figure 1, 112 studies and 149 effect sizes were finally included in the meta-analysis. The studies included in the meta-analysis could not be presented in the article due to page limitations. Therefore, they are shown in the original thesis mentioned in the footnote on the article's first page. Those who want to access the primary studies can access the thesis presented in the references.

Coding of Data

The coding of the 112 studies included in the meta-analysis and the 149 effect sizes obtained from these studies were made in Microsoft Excel. The descriptive variables considered in the coding made in Microsoft Excel are: "publication code, name of the study, colophon (author surnames, year of publication), publication type, publication language, publication year, place of publication, volume-number, authors, database, index, models used, additional application in the treatment group, techniques used in the control group, research design, subject area, grade level, course area (science, physics, chemistry, biology), data collection tool, dependent variable, reliability coefficient, range of difficulty/mean difficulty, population-sample, number of activities, class hours, piloting status (yes/no), the piloting status of achievement test (yes/no), data analysis method, application time, school type". The categorical variables determined for coding and the number of studies and effect sizes in the categories of these variables are given in Table 2.

Table 2

Number of studies and effect sizes for coded categorical variables

	Number of studies (f)	Number of effect sizes (f)		Number of studies (f)	Number of effect sizes (f)
Study Type			**Study Language**		
Article	48	55	English	23	26
Proceeding	9	11	Turkish	89	123
Master’s Thesis	37	45	**Databases**		
Doctoral Thesis	18	38	Google Scholar	67	80
Publishing Time			Dergipark	2	2
2005-2009	25	32	ERIC	6	7
2010-2014	46	69	Taylor & Francis	1	2
2015-2020	41	48	HEI Thesis Center	27	48
Study Design			Science Direct	5	5
True experimental	3*	3	Web of Science	3	4
Quasi experimental	96*	124	Proquest	1	1
Poor experimental	14	22			
Grade Level			**Subject**		
4. and 5.	10	13	Science	1	1
6., 7. and 8.	50	63	Physic	44	57
9.,10., 11. and 12.	36*	51	Chemistry	36	52
High education	18*	22	Biology	31	39
Academic Year			**School Type**		
Unspecified	14	20	Unspecified	4	4
(2001-2002)-(2007-2008)	28	36	Public	102	134
(2008-2009)-(2013-2014)	46	66	Private	5	10
(2014-2015)-(2019-2020)	24	27	Public and Private	1	1
Total	112	149		112	149

*One of the studies used both true experimental design and quasi-experimental design.

The statistics related to effect sizes were also coded in the same file for performing the meta-analysis study. Since some primary studies reported effect sizes directly, Cohen *d*, Hedges *g,* and *η2* effect sizes were taken directly, and the sample size of the treatment groups and the sample size of the control group were also coded. In addition, in some primary studies, the statistics required to calculate effect sizes were coded, and thus effect sizes were calculated. For the true and quasi-experimental designs that calculated statistics such as mean and standard deviation, the mean and standard deviation for the post-test of the treatment group and the mean and standard deviations for the post-test of the control group were coded. If the research was conducted in a weak experimental design, the means and standard deviations for both the post-test and pre-test of the treatment group were included in the coding. In addition, if mean and standard deviation values were not reported in the studies that also used analyses such as *t*-test, ANOVA, Mann Whitney U Test, ANCOVA, MANOVA, MANCOVA, Wilcoxon Signed-Rank Test, and Kruskal-Wallis H Test, statistics related to these analyses were coded, and effect sizes were calculated according to these statistics. Finally, correlation was coded for primary studies that reported correlation coefficient as correlation directly means effect size.

Data Analysis

In the meta-analysis examining the effect of the 5E teaching method on academic achievement in science education, it was examined how the overall effect sizes were affected when weighting with the inverse of the sampling error variance and reliability were applied. In addition, it was also examined how the overall effect sizes were affected when gray literature was included and was not included. CMA program and random effects model were used to obtain the overall effect sizes. Two different types of weighting were used in the CMA program. The first one is weighting by the inverse of the sampling error variance (Hedges & Vevea, 1998), and how it is calculated is shown in Equation 1 (Borenstein et al., 2009);

$$w_i^* = \frac{1}{V_{yi}^*} \tag{1}$$

In Equation 1, w_i* represents the weight of the relevant study for the random effects model, while V_{yi}* is the sum of the sampling error variance (*Vyi*) of the relevant study to be weighted and the variance between studies (T^2). For weighting by reliability coefficient, the weighting is as in Equation 2 for fixed effects and random effects models. However, within the scope of the research, meta-analysis was conducted according to the random effects model.

$$w_i = r_{at} \qquad w_i^* = r_{at} + T^2 \tag{2}$$

In Equation 2, while w_i represents the weight of the related study, r_{at} represents the reliability coefficient for the measurements obtained with the achievement test used in the related study. T^2 represents the variance between studies and is used to calculate w_i^* in the random effects model.

The weighting types determined were used both for the cases where gray literature was included in the meta-analysis and for the cases where it was not included, and the overall effect sizes and standard errors obtained were interpreted. There were 149 effect sizes in the meta-analysis when gray literature was included, while there were 55 effect sizes when gray literature was excluded. In addition to interpreting the effect of the inclusion and exclusion of gray literature on the meta-analysis results, it was examined whether there was a significant difference between the effect sizes between the studies in the gray literature and the articles. Accordingly, a Q test based on analysis of variance was performed.

Before conducting the meta-analyses, the heterogeneity values for the data were examined with *Q, p(Q),* T^2, I^2, H^2 and R^2 statistics. For the I^2 statistic, 25% is interpreted as low, 50% as medium and 75% as high heterogeneity (Higgins et al., 2003). H^2 and R^2 statistics of 1 is an indication of homogeneity of effect sizes. Publication bias was examined with the funnel plot and trim-and-fill method by Duval and Tweedie (Duval & Tweedie, 2000a; 2000b), Rosenthal's fail-safe *N*, Begg and Mazumdar's rank correlation test and Egger's regression intercept methods. The number of missing studies calculated in

Rosenthal's fail-safe *N* method was compared with the criterion value of 5*k*+10 (k=number of studies) (Rosenthal, 1979). In Begg and Mazumdar's rank correlation and Egger's regression intercept methods, the significance of the correlation and intercept were interpreted, respectively (Begg & Mazumdar, 1994; Egger et al., 1997).

Results

Heterogeneity

Within the scope of the study, firstly, heterogeneity and publication bias regarding the primary studies included in the meta-analysis were examined. The heterogeneity statistics, *Q, p(Q),* T^2, I^2, H^2 *ve* R^2, were analyzed and given in Table 3.

Table 3

Heterogeneity statistics

k	*Q*	*df*	*p*	T^2	I^2	H^2	R^2
149	1102.69	148	0.000*	0.455	%86.578	7.450	7.796

*p < .001

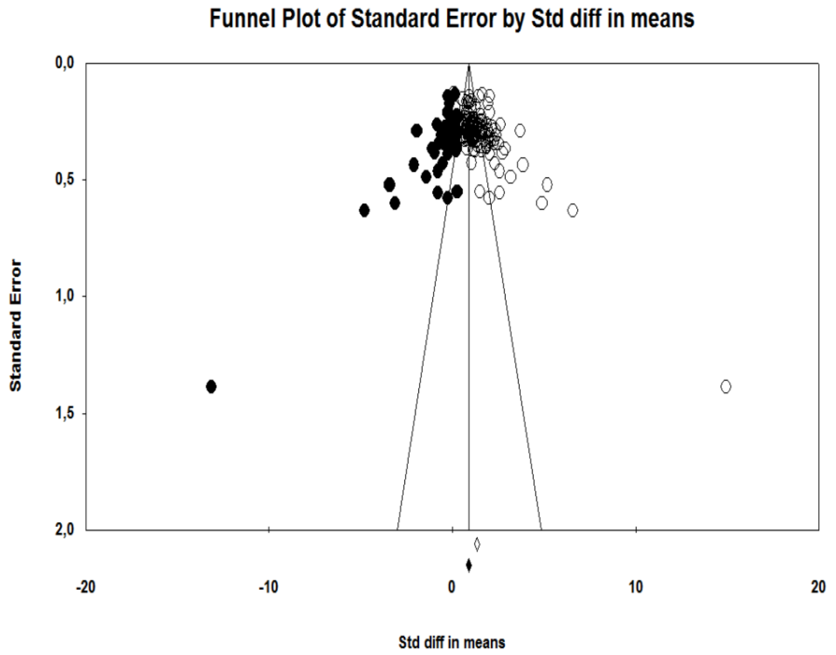
When Table 3 is analyzed, it is seen that the *Q* value is significant. While this is an indicator of heterogeneity, an I^2 value higher than 75% is an indicator of high heterogeneity (Higgins et al., 2003). Besides, the fact that the T^2 value is quite different from 0 indicates the presence of variance between studies. In addition, the fact that H^2 and R^2 statistics are quite different from 1 indicates that effect sizes are heterogeneously distributed (Higgins & Thompson, 2002). When all statistics are handled together, it is observed that heterogeneity exists. In addition to statistical evidence, there is also theoretical evidence for the existence of heterogeneity. The fact that the studies included in the meta-analysis belong to different populations is also a source of heterogeneity. For example, the research data has a wide range of education levels from secondary school to higher education. Furthermore, the regions where the primary studies were conducted differ from each other in many aspects, such as climate and culture. Moreover, the subject areas in the primary studies differ from each other in physics, chemistry, biology, and science. Based on this, when the statistical and theoretical evidence of heterogeneity is considered together, it can be said that the weighting methods in this study were compared under a condition where heterogeneity exists.

Publication Bias

The study analyzed publication bias using the funnel plot and Duval and Tweedie’s trim-and-fill method, Rosenthal's fail-safe N method, Begg and Mazumdar's rank correlation, and Egger's regression intercept method. The funnel plot is given in Figure 2.

Figure 2

Funnel Plot



The funnel diagram in Figure 2 shows that studies (filled dots) had to be added to adjust the symmetry of the plot. This indicates publication bias and the diagram is evaluated together with Duval and Tweedie's trim-and-fill results in Table 4.

Table 4

The results of Duval & Tweedie's trim-and-fill

	Studies Trimmed	Overall Effect	Lower Limit	Upper Limit	Q Value
Observed Values		1.347	1.228	1.466	1102.690
Adjusted Values	48	0.912	0.777	1.046	2379.926

In Table 4, it was observed that 48 studies were added to make the funnel plot symmetrical and the added studies changed the overall effect. In addition, in Rosenthal's fail-safe *N* method, it was observed that the number of missing studies that should be added for the overall effect size to be non-significant was 177019, and this value was greater than the criterion value of 755 (5k+10) (Rosenthal, 1979). When Begg and Mazumdar's rank correlation results were analyzed, it was seen that Kendall's tau value was 0.326 and significant. Finally, in Egger's regression intercept method, the intercept was found to be 3.834 and significant. The fact that these statistics are significant is an indicator of publication bias. When all statistics are evaluated together, it is observed that there is publication bias. Based on this, it can be said that the weighting methods in this study were compared under a condition where publication bias exists.

Meta-Analysis Results

In this study, the effect of weighting with the inverse of the sampling error variance and reliability in the presence of high heterogeneity and publication bias on meta-analysis results was examined. We also examined the effect of the inclusion and exclusion of gray literature on the meta-analysis results and the results are presented in Table 5.

Table 5

The results of meta-analysis in different conditions (weighting methods and gray literature)

Weighting Methods	Number of Effect Sizes	Cohen *d*	SE	Variance	Lower Limit	Upper Limit	Z	p
Gray Literature Included								
Inverse variance	149	1.347	0.061	0.004	1.228	1.466	22.217	0.000
Reliability	149	1.474	0.119	0.014	1.242	1.707	12.426	0.000
Gray Literature Excluded								
Inverse variance	55	1.281	0.076	0.006	1.132	1.431	16.780	0.000
Reliability	55	1.324	0.152	0.023	1.026	1.622	8.705	0.000

When Table 5 was examined, it was seen that the largest overall effect size was obtained in the weighting method with a reliability of 1.474, and the smallest overall effect size was obtained in the weighting method with the inverse of sampling error variance with 1.347 when gray literature is included. When the standard error values were analyzed, it was seen that the lowest standard error value was obtained from weighting with a sampling error variance of 0.061. The highest standard error value was found in weighting by reliability coefficient, which was 0.119. Variance values also changed in parallel with the standard error values. When evaluated in terms of confidence interval, the narrowest confidence interval was found in the sampling error variance method, again in parallel with the standard error. In addition, the confidence interval was wider for the weighting method with the reliability coefficient. When the significance of the overall effect sizes was analyzed, it was observed that the overall effect sizes were significant in both methods. In addition, forest plots of both methods are presented in Appendix A and Appendix B, respectively. When the forest plots were analyzed, it was seen that the primary studies were more homogeneous in terms of confidence intervals due to the narrow range of weights in the reliability weighting method. On the other hand, when the weighting method with sampling error variance was used, it could be said that the forest plot was more heterogeneous due to the wide sample range.

In the case where gray literature was not included, the largest overall effect size was obtained from weighting methods with a reliability coefficient and was found to be 1.324. The lowest overall effect size was found to be 1.281 for the weighting by sampling error variance method. When the standard error values were analyzed, it was seen that the lowest standard error value was obtained from weighting with sampling error variance and was 0.076. The highest standard error value was found in weighting by reliability coefficient, which was 0.152. Variance values also changed in parallel with the standard error values. When the confidence intervals were evaluated, it could be said that the confidence interval was wider when weighting by reliability coefficient than when weighting by sampling error variance. It was observed that the meta-analysis study with the narrowest confidence interval was the meta-analysis using the weighting method with sampling error variance. When the significance of the overall effect sizes was analyzed, it was seen that the overall effect sizes were significant in both methods.

In addition to interpreting the effects of the inclusion and exclusion of gray literature on the meta-analysis results, it is also necessary to interpret the significance of these effects. In this context, Analog ANOVA was conducted to examine the significance of the effects. The results are given in Table 6.

Table 6

Analog ANOVA results of gray literature and articles for weighting methods

Weighting Method		**Q values**	**df (Q)**	**p**
Inverse Variance	Within Group	1101.062	147	0.000
$N_{GrayLiterature}$ = 94	Between Groups	1.629	1	0.202
$N_{Manuscript}$ = 55	Total	1102.690	148	0.000
Reliability	Within Group	240.668	147	0.000
$N_{GrayLiterature}$ = 94	Between Groups	1.597	1	0.206
$N_{Manuscript}$ = 55	Total	242.265	148	0.000

When Table 6 was examined, it was seen that the p-values for the intergroup Q values in the inverse of the sampling error variance and reliability weighting methods were 0.202 and 0.206, respectively. In this respect, it was clear that the difference between the average effect size obtained from the studies in the gray literature and the average effect size obtained from the articles was not significant in all weighting methods. Therefore, it can be said that the meta-analysis results obtained with and without the inclusion of gray literature did not differ significantly from each other.

Discussion

When the weighting methods were compared with each other, both when gray literature was included and not included in the meta-analyses, it was seen that the weighting method with the smallest overall effect size was the weighting method with the sampling error variance. The weighting method with the largest overall effect size was the weighting method with a reliability coefficient. The fact that the overall effect size obtained from weighting with sampling error variance is lower than the effect sizes obtained from weighting with reliability coefficient does not indicate that the weighting method with reliability coefficient synthesizes effect size more accurately than the weighting method with sampling error variance. The reason for the difference in the overall effect sizes between the two weighting methods may be that weighting by sampling error variance deals with the sampling error, whereas weighting by reliability coefficient deals not only with sampling error but also with sources of random error, including sampling error. In addition, the fact that the overall effect sizes are larger in the weighting method with reliability coefficient may be due to the fact that, as Rosenthal (1991) states, the contribution of studies that are weaker in terms of quality weight and have smaller effect sizes to the average effect size is less than other studies.

A similar situation is observed when standard error values are examined in the context of weighting methods. It was observed that the standard error values obtained from weighting by reliability coefficient were the highest, while the standard error values obtained from weighting by sampling error variance were the lowest, both in the conditions where gray literature was included and not included. The fact that the standard error values obtained from weighting with sampling error variance were lower than the standard error values obtained from weighting with reliability coefficient can be explained by the fact that it deals only with the dimension of the error arising from the sample. This is because weighting with the reliability coefficient addresses not only sampling error but also other sources of random error sources. Therefore, the standard error values obtained from the weighting methods with sampling error variance and reliability coefficient differ from each other. In parallel with the standard error, the narrowest confidence intervals were observed in the weighting method with sampling error variance in all studies, while the widest confidence intervals were observed in the weighting method with reliability coefficient. This is because the confidence interval is calculated directly using the standard error. The fact that the lower and upper limit values obtained from weighting with sampling error variance are lower than the other lower and upper limit values and the confidence intervals are narrower can be explained by the fact that only the error arising from the sample is considered in parallel with the overall effect size and standard error.

When the meta-analysis results were compared according to the inclusion and exclusion of gray literature, it was observed that the overall effect size had different values and the overall effect sizes were higher when the gray literature was included. However, it was concluded that this difference was not significant in both weighting methods. Although the difference was not significant, the reason why the overall effect sizes were higher when the gray literature was included might be due to the fact that the effect sizes of the primary studies in the gray literature were larger than the scanned studies. In addition, higher average effect sizes may have been obtained due to the larger sample sizes of these studies where effect sizes might be larger.

Like the overall effect size, the standard error also took different values according to the inclusion of gray literature. In general, standard error values were higher when gray literature was not included. The standard error is expected to decrease as the sample size increases with the inclusion of gray literature. Conn et al. (2003) stated that when gray literature was included, the overall effect size was estimated with less error than when gray literature was not included, which is similar to the results of weighting with sampling error variance and reliability coefficient in this study. Moher et al. (1996), similar to the results of this study, found that there was a slight difference due to the reporting language of the studies but that this was not a significant bias and that the inclusion of non-English language publications may reduce the error and increase the accuracy of estimation. As stated by Conn et al. (2003) and Moher (1996) in their studies and as found in this study, the reason for the decrease in the standard error and more accurate estimations may be the increase in the number of included studies. Hartling et al. (2017) have also observed that the studies included in the gray literature generally constitute a very small part of the meta-analysis sample, and therefore, the results are not affected much by the inclusion of the gray literature. However, in this study, the studies in the gray literature constitute a larger portion of the studies rather than a small portion of the studies. Despite this, the effect of the inclusion of gray literature is not significant and is similar to Hartling et al. (2017). Contrary to the results of this study, Corlett (2011) also stated that ignoring the gray literature might lead to biased results. Although Corlett (2011) did not statistically examine the effect of gray literature, the reason why he made such a suggestion is that he worked in the tropics and gray literature is the only source in the tropics. Based on the findings of this study and the literature, it is obvious that it is important to investigate the impact of gray studies in order to make a correct decision about whether there is bias in a meta-analysis study.

Conclusions, Suggestions and Limitations

The study results showed that the overall effect size changed with the inverse of the sampling error variance and when weighted by reliability. It was also concluded that the standard error was highest when weighted by the reliability coefficient because it included all random errors. In this regard, meta-analysts may also be recommended to try weighting with a reliability coefficient because it is thought that weighting by reliability may provide a more accurate confidence interval.

When the results regarding the inclusion of gray literature were examined, it was observed that the results were not significantly different in the inclusion and exclusion cases. In this study, although there was no significant difference between the overall effect sizes according to the inclusion of gray literature, it is recommended that researchers should also scan the gray literature in all weighting methods since the estimation accuracy will increase due to the lower standard error when gray literature is included.

When the weighting methods were compared with each other, it was seen that weighting with sampling error variance gave the closest results when gray literature was included and not included. Therefore, when weighting with sampling error variance, the exclusion of gray literature may be less important for educational research. However, since clinical research requires more precise results, it may be recommended to include the gray literature since these studies have little differentiation. The weighting method with the highest differentiation was found to be weighting with a reliability coefficient. Although this differentiation is not significant, there may be significance between the inclusion and exclusion of gray literature in other studies. For this reason, researchers are strongly recommended to review the gray literature and examine the significance of the difference when using weighting with a reliability coefficient.

Within the scope of this study, the results were compared with each other by weighting with reliability coefficient in addition to weighting with sampling error variance used in classical meta-analysis. Other researchers can compare meta-analysis results by formulating different weighting methods or choosing not to weight. They can also contribute to the mathematical formulation of the weighting method with reliability. In addition, other researchers can choose another study topic instead of the effect of the 5E teaching model on science achievement, which was selected as the subject of the meta-analysis study in this study, or they can compare the methods in this study in fields such as sports sciences, health sciences, etc. instead of using data in the field of education.

In the present study, there is a situation of publication bias and high heterogeneity, which are the limitations of the study. Other researchers can examine the method of weighting effect sizes with the reliability coefficient developed in this study under different conditions. For this purpose, they can design a simulation study and test this new method under conditions of different sample sizes, number of studies, estimation methods, heterogeneity, publication bias, fields, etc. As a result, this study is expected to encourage new studies on weighting the measures from which effect sizes are obtained with reliability coefficients in synthesizing studies in meta-analysis and to add the options of the reliability of measures for weighting effect sizes to meta-analysis softwares.

Declarations

Conflict of Interest: No potential conflict of interest was reported by the authors.
Ethical Approval: Permission was obtained from Gazi University Ethics Commission dated 22.02.2021 and numbered E-77082166-302.08.01-33880.
Author Contribution: Yıldız YILDIRIM: conceptualization, investigation, methodology, writing - original draft, formal analysis, visualization, editing. Şeref TAN: conceptualization, investigation, methodology, data curation, supervision, writing - review & editing

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Consent to Participate: All authors have given their consent to participate in submitting this manuscript to this journal.

Consent to Publish: Written consent was sought from each author to publish the manuscript.

References

Alacapınar, F. G., & Ok, M. (2020). Meta-analysis covering studies on problem-based learning. *Research on Education and Psychology*, *4*(Special Issue), 53-73. https://dergipark.org.tr/tr/pub/rep/issue/54042/703272

Altunoğlu, B.D., Bozdemir Yüzbaşıoğlu, H., Candan Helvacı, S., & Kurnaz, M.A. (2020). Genetik kavramlara ilişkin eğitim çalışmalarının meta analiz yöntemi ile incelenmesi. *Batı Anadolu Eğitim Bilimleri Dergisi, 11*(2), 643-661. https://dergipark.org.tr/en/pub/baebd/issue/58594/702868

Arık, S., & Yılmaz, M. (2020). The effect of constructivist learning approach and active learning on environmental education: A meta-analysis study. *International Electronic Journal of Environmental Education, 10*(1), 44-84. https://dergipark.org.tr/tr/pub/iejeegreen/issue/49969/605746

Basit, O. (2020). *Türkiye'de yapılan okul öncesi dönem çocuklarının gelişim alanlarını destekleyici çalışmaların incelenmesi: Bir meta analiz çalışması.* Doktora Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Başpınar, M. M. (2021). Türkiye'de yapılan tez çalışmalarında sigara içiminin uyku kalitesi üzerine etkisinin değerlendirilmesi: meta-analiz. *Journal of Turkish Sleep Medicine*, *8*(1), 7-15. https://doi.org/10.4274/jtsm.galenos.2021.98698

Batdı, V., & Batdı, H. (2015). Effect of creative drama on academic achievement: A meta-analytic and thematic analysis. *Educational Sciences: Theory & Practice, 15*(6), 1459-1470. https://doi.org/10.12738/estp.2015.6.0156

Becker, K. H., & Park, K. (2011). Integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A meta-analysis. *Journal of STEM Education Innovation and Research, 12*(5), 23-37. https://doi.org/10.12691/education-2-10-4

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088-1101. PMID: 7786990. https://doi.org/10.2307/2533446

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

Bozdemir, H., Çevik, E. E., Altunoğlu, B. D., & Kurnaz, M. A. (2017). Astronomi konularının öğretiminde kullanılan farklı yöntemlerin akademik başarıya etkisi: Bir meta analiz çalışması. *Alan Eğitimi Araştırmaları Dergisi*, *3*(1), 12-24. https://dergipark.org.tr/tr/download/article-file/266427

Conn, V. S., Valentine, J. C., Cooper, H. M., & Rantz, M. J. (2003). Grey literature in meta analyses. *Nursing Research*, 52, 256–261. https://doi.org/10.1097/00006199-200307000-00008

Corlett, R. T. (2011). Trouble with the gray literature. *Biotropica*, *43*(1), 3-5. https://doi.org/10.1111/j.1744-7429.2010.00714.x

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

Duval, S., & Tweedie, R. (2000b). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463. https://doi.org/10.1111/j.0006-341x.2000.00455.x.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

Englund, G., Sarnelle, O., & Cooper, S. D. (1999). The importance of data-selection criteria: meta-analyses of stream predation experiments. *Ecology*, *80*(4), 1132-1141. https://www.jstor.org/stable/177060?seq=1#metadata_info_tab_contents

Fabiano, G., Schatz, N., Aloe, A., Pelham, W., Smyth, A., Zhao, X., Merrill, B. M., Macphee, F., Ramos, M., Hong, N., Altszuler, A., Ward, L., Rodgers, D. B., Liu, Z., Karatoprak Ersen, R., & Coxe, S. (2021). Comprehensive meta-analysis of attention-deficit hyperactivity disorder psychosocial treatments investigated within between group studies. Review of Educational Research, 91(5), 718–760. https://doi.org/10.3102/00346543211025092

Fuller, J. B., & Hester, K. (1999). Comparing the sample-weighted and unweighted meta-analysis: An applied perspective. Journal of Management, 25(6), 803-828. https://doi.org/10.1177/014920639902500602

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3-8. https://doi.org/10.2307/1174772

Hartling, L., Featherstone, R., Nuspl, M., Shave, K., Dryden, D. M., & Vandermeer, B. (2017). Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. *BMC Medical Research Methodology*, *17*(1), 1-11. https://doi.org/10.1186/s12874-017-0347-z

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

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486–504. https://doi.org/10.1037/1082-989X.3.4.486

Higgins, J. P., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions.* The Cochrane Collaboration and John Wiley & Sons.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta analysis. *Statistics in Medicine, 21*(11), 1539-1558. https://doi.org/10.1002/sim.1186

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, *327*(7414), 557. https://doi.org/10.1136/bmj.327.7414.557

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Beverly Hills, CA: Sage.

Hunter, J. E., & Schmidt, F. L. (2004). Methods of meta-analysis: Correcting error and bias in research findings (2nd ed.). Thousand Oaks, CA: Sage.

Karasar, N. (2013). *Bilimsel araştırma yöntemleri* (25th ed.). Nobel.

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, *86*(3), 681-718. https://doi.org/10.3102/0034654315627366

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology*, *62*(10), 1-34. https://doi.org/10.1136/bmj.b2700

Marín-Martínez, F., & Sánchez-Meca, J. (2009). Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational and Psychological Measurement*, *70*(1), 56-73. https://doi.org/10.1177/0013164409344534

Moher, D., Fortin, P., Jadad, A. R., Jüni, P., Klassen, T., Le Lorier, J., ... & Linde, K. (1996). Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *The Lancet*, *347*(8998), 363-366. https://doi.org/10.1016/s0140-6736(96)90538-3

Mutluer, C., Gündüz, T., Çelikten Demirel, S., & Çakan, M. (2020). *Meta analiz çalışmasında sabit etki modeline karşı rastgele etki modeli.* Presented at the VIIth International Eurasian Educational Research Congress, Eskişehir.

National Research Council. (1992). *Combining information: statistical issues and opportunities for research*. National Academies. https://doi.org/10.17226/20865

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). Quality-weighting of studies in meta-analytic research. *Psychotherapy Research, 1*(1), 25-28. https://doi.org/10.1080/10503309112331334031

Saraç, H. (2018). Yapılandırmacı yaklaşım öğrenme halkası modellerinin öğrenilen bilgilerin kalıcılığına etkisi: Meta analiz çalışması. *Kastamonu Eğitim Dergisi*, *26*(3), 753-764. https://doi.org/10.24106/kefdergi.413322

Schmidt, F. L., Oh, I. S., & Hayes, T. L. (2009). Fixed-versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, *62*(1), 97-128. https://doi.org/10.1348/000711007X255327

Shuster, J. J. (2010). Empirical vs natural weighting in random effects meta-analysis. *Statistics in Medicine*, *29*(12), 1259-1265. https://doi.org/10.1002/sim.3607

Simpson, R. J. S., & Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *The British Medical Journal*, *2*(2288), 1243-1246. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2355479/

Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research*, *81*(1), 97-128. https://doi.org/10.3102/0034654310378174

Toraman, Ç. , Çelik, Ö. C. & Çakmak, M. (2018). Oyun-Tabanlı Öğrenme Ortamlarının Akademik Başarıya Etkisi: Bir Meta–Analiz Çalışması . *Kastamonu Eğitim Dergisi* , *26* (6) , 1803-1811 . https://doi.org/10.24106/kefdergi.2074

Özdemir, V. (2023). *Okuma becerisi ile sosyoekonomik ve kültürel değişkenler arasındaki ilişkilerin incelenmesi meta-analiz çalışması*. Presented at the X International Eurasian Educational Research Congress, Ankara.

Warfa, A. R. M. (2016). Using cooperative learning to teach chemistry: A meta-analytic review. *Journal of Chemical Education, 93*(2), 248-255. https://doi.org/10.1021/acs.jchemed.5b00608

Xie C., Wang, M., & Hu, H. (2018). Effects of constructivist and transmission instructional models on mathematics achievement in mainland China: A meta-analysis. *Front. Psychol.,* *9*(1923), 1-18. https://doi.org/10.3389/fpsyg.2018.01923

Yeşilpınar Uyar, M. & Doğanay, A. (2018). Öğrenci merkezli strateji, yöntem ve tekniklerin akademik başarıya etkisi: Bir meta-analiz çalışması. *Mersin Üniversitesi Eğitim Fakültesi Dergisi, 14*(1), 186-209. https://doi.org/10.17860/mersinefd.334542

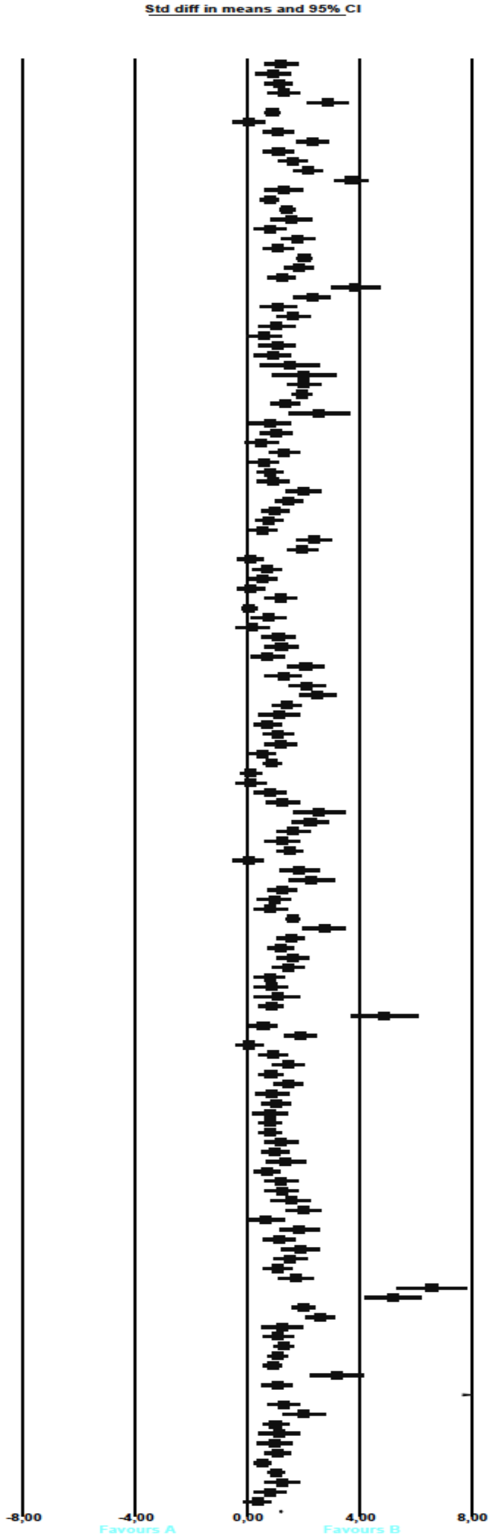
Yıldırım, Y., & Şahin, M. G. (2023). How Do Different Weighting Methods Affect the Overall Effect Size in Meta-Analysis?: An Example of Science Attitude in Türkiye Sample. *International Journal of Psychology and Educational Studies, 10*(3), 744-757. https://doi.org/10.52380/ijpes.2023.10.3.1049

APPENDIX

APPENDIX A

The forest plot for inverse variance method

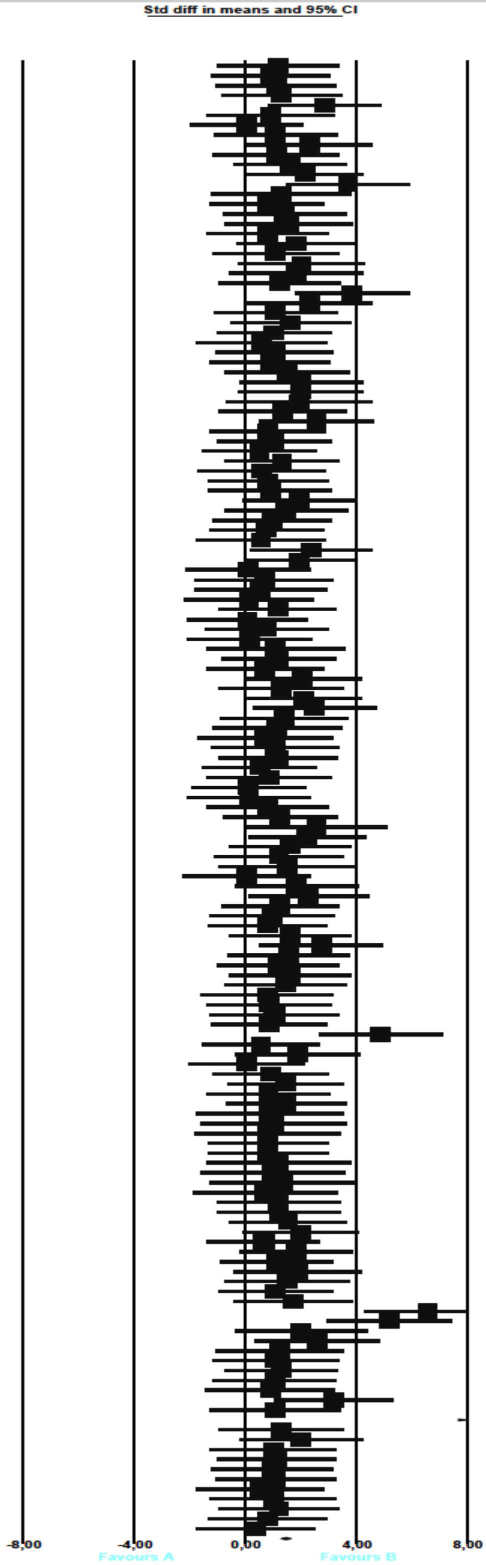
Study name	Statistics for each study						
	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
1001,000	1,200	0,307	0,094	0,598	1,802	3,906	0,000
1003,000	0,921	0,317	0,101	0,300	1,543	2,905	0,004
1004,000	1,127	0,256	0,065	0,626	1,628	4,406	0,000
1005,000	1,312	0,285	0,081	0,754	1,870	4,609	0,000
1006,000	2,878	0,368	0,136	2,156	3,600	7,814	0,000
1007,000	0,907	0,144	0,021	0,624	1,190	6,287	0,000
1008,000	0,064	0,279	0,078	-0,484	0,611	0,227	0,820
1009,000	1,109	0,277	0,077	0,565	1,652	3,997	0,000
1010,000	2,326	0,286	0,082	1,765	2,886	8,134	0,000
1012,000	1,119	0,285	0,081	0,560	1,678	3,923	0,000
1015-1	1,629	0,252	0,064	1,135	2,123	6,463	0,000
1015-2	2,179	0,276	0,076	1,639	2,720	7,905	0,000
1016,000	3,712	0,292	0,085	3,140	4,284	12,726	0,000
1017,000	1,306	0,337	0,113	0,646	1,966	3,880	0,000
1020,000	0,797	0,164	0,027	0,475	1,119	4,849	0,000
1022,000	1,431	0,142	0,020	1,152	1,710	10,058	0,000
1023,000	1,578	0,377	0,142	0,839	2,316	4,189	0,000
1024,000	0,803	0,294	0,087	0,226	1,379	2,728	0,006
1026-1	1,821	0,303	0,092	1,228	2,414	6,019	0,000
1026-2	1,105	0,273	0,075	0,570	1,641	4,046	0,000
1028,000	2,033	0,142	0,020	1,755	2,312	14,298	0,000
1031,000	1,860	0,261	0,068	1,348	2,372	7,121	0,000
1034,000	1,240	0,250	0,063	0,749	1,731	4,949	0,000
1035,000	3,862	0,437	0,191	3,006	4,718	8,838	0,000
1036,000	2,315	0,328	0,108	1,671	2,958	7,050	0,000
1039,000	1,101	0,331	0,110	0,452	1,750	3,325	0,001
1040,000	1,650	0,299	0,089	1,065	2,236	5,521	0,000
1041-1	1,058	0,330	0,109	0,412	1,704	3,208	0,001
1041-2	0,613	0,316	0,100	-0,007	1,233	1,939	0,053
1041-3	1,075	0,330	0,109	0,428	1,723	3,254	0,001
1041-4	0,910	0,325	0,105	0,274	1,546	2,805	0,005
1042-1	1,517	0,551	0,304	0,437	2,598	2,753	0,006
1042-2	2,022	0,579	0,336	0,886	3,158	3,490	0,000
1042-3	2,016	0,307	0,094	1,413	2,618	6,558	0,000
1043,000	1,940	0,175	0,031	1,597	2,283	11,080	0,000
1048,000	1,369	0,266	0,071	0,849	1,890	5,156	0,000
1049,000	2,573	0,555	0,308	1,486	3,661	4,637	0,000
1051,000	0,813	0,368	0,135	0,092	1,534	2,210	0,027
1052-1	1,051	0,293	0,086	0,475	1,626	3,580	0,000
1052-2	0,518	0,314	0,099	-0,098	1,133	1,649	0,099
1052-3	1,331	0,274	0,075	0,793	1,869	4,851	0,000
1054,000	0,588	0,273	0,075	0,053	1,123	2,155	0,031
1055,000	0,818	0,237	0,056	0,353	1,284	3,446	0,001
1056,000	0,916	0,292	0,085	0,345	1,488	3,141	0,002
1059,000	1,980	0,321	0,103	1,352	2,608	6,177	0,000
1061,000	1,478	0,252	0,064	0,983	1,972	5,857	0,000
1065-1	0,997	0,253	0,064	0,501	1,494	3,935	0,000
1065-2	0,786	0,248	0,062	0,300	1,272	3,168	0,002
1067,000	0,574	0,263	0,069	0,058	1,091	2,180	0,029
1072-1	2,386	0,311	0,096	1,777	2,995	7,683	0,000
1072-2	1,966	0,289	0,084	1,399	2,532	6,800	0,000
1076,000	0,117	0,239	0,057	-0,352	0,586	0,490	0,624
1077,000	0,686	0,266	0,071	0,166	1,207	2,583	0,010
1079,000	0,560	0,257	0,066	0,056	1,063	2,180	0,029
1080,000	0,141	0,245	0,060	-0,339	0,620	0,575	0,565
1081,000	1,192	0,305	0,093	0,594	1,790	3,906	0,000
1082,000	0,085	0,133	0,018	-0,174	0,345	0,644	0,520
1084-1	0,767	0,314	0,098	0,152	1,382	2,446	0,014
1084-2	0,171	0,306	0,094	-0,429	0,772	0,559	0,576
1087,000	1,111	0,305	0,093	0,514	1,708	3,646	0,000
1088,000	1,217	0,318	0,101	0,595	1,840	3,832	0,000
1089,000	0,728	0,315	0,099	0,110	1,345	2,310	0,021
1092,000	2,087	0,341	0,117	1,417	2,756	6,112	0,000
1093,000	1,296	0,332	0,110	0,646	1,946	3,908	0,000
1094,000	2,130	0,330	0,109	1,483	2,777	6,454	0,000
1095,000	2,516	0,346	0,119	1,839	3,193	7,281	0,000
1099,000	1,406	0,266	0,071	0,884	1,928	5,276	0,000
1102,000	1,162	0,374	0,140	0,430	1,894	3,110	0,002
1103,000	0,728	0,253	0,064	0,233	1,224	2,879	0,004
1104,000	1,107	0,273	0,074	0,572	1,642	4,057	0,000
1110,000	1,200	0,280	0,079	0,650	1,750	4,279	0,000
1113,000	0,535	0,252	0,064	0,040	1,029	2,117	0,034
1115,000	0,875	0,171	0,029	0,540	1,211	5,114	0,000
1116,000	0,129	0,193	0,037	-0,249	0,507	0,670	0,503
1117-1	0,138	0,280	0,079	-0,411	0,688	0,494	0,621
1117-2	0,808	0,289	0,083	0,242	1,374	2,799	0,005
1118,000	1,271	0,310	0,096	0,663	1,879	4,099	0,000
1126,000	2,576	0,465	0,216	1,665	3,487	5,544	0,000
1127,000	2,252	0,343	0,118	1,579	2,925	6,560	0,000
1903-E	1,650	0,299	0,089	1,065	2,236	5,521	0,000
1301,000	1,233	0,322	0,104	0,602	1,864	3,829	0,000
1303,000	1,529	0,227	0,051	1,085	1,974	6,747	0,000
1401,000	0,050	0,272	0,074	-0,484	0,583	0,182	0,855
1405,000	1,857	0,358	0,128	1,155	2,559	5,184	0,000
1406,000	2,298	0,429	0,184	1,456	3,140	5,350	0,000
1409,000	1,255	0,262	0,068	0,742	1,767	4,797	0,000
1410,000	0,969	0,304	0,093	0,372	1,565	3,181	0,001
1412,000	0,819	0,307	0,094	0,218	1,421	2,669	0,008
1413-1	1,627	0,135	0,018	1,362	1,893	12,030	0,000
1413-2	2,744	0,386	0,149	1,987	3,500	7,111	0,000
1413-3	1,573	0,249	0,062	1,085	2,062	6,311	0,000
1413-4	1,203	0,240	0,057	0,733	1,673	5,019	0,000
1413-5	1,642	0,285	0,081	1,084	2,199	5,770	0,000
1414,000	1,475	0,291	0,085	0,905	2,046	5,066	0,000
1416,000	0,807	0,279	0,078	0,261	1,353	2,896	0,004
1422,000	0,868	0,302	0,091	0,276	1,460	2,875	0,004
1423,000	1,065	0,428	0,183	0,226	1,903	2,489	0,013
1424,000	0,861	0,221	0,049	0,428	1,295	3,892	0,000
1425,000	4,893	0,603	0,363	3,713	6,074	8,122	0,000
1428,000	0,575	0,259	0,067	0,067	1,083	2,218	0,027
1432-1	1,882	0,292	0,085	1,309	2,454	6,442	0,000
1433-1	0,070	0,252	0,064	-0,425	0,564	0,277	0,782
1433-2	0,923	0,261	0,068	0,411	1,435	3,535	0,000
1433-3	1,459	0,301	0,090	0,869	2,048	4,851	0,000
1435,000	0,849	0,218	0,047	0,423	1,276	3,900	0,000
1436,000	1,478	0,266	0,071	0,957	1,999	5,557	0,000
1437-1	0,896	0,300	0,090	0,308	1,484	2,985	0,003
1437-2	1,034	0,265	0,070	0,515	1,553	3,903	0,000
1437-3	0,813	0,325	0,106	0,176	1,450	2,502	0,012
1438-1	0,823	0,211	0,045	0,409	1,237	3,895	0,000
1438-2	0,823	0,211	0,045	0,409	1,237	3,895	0,000
1439-1	1,215	0,311	0,097	0,605	1,825	3,907	0,000
1439-2	1,008	0,258	0,067	0,502	1,515	3,903	0,000
1439-3	1,376	0,352	0,124	0,686	2,066	3,909	0,000
1439-4	0,723	0,233	0,054	0,266	1,180	3,101	0,002
1440-1	1,215	0,311	0,097	0,605	1,825	3,907	0,000
1440-2	1,230	0,315	0,099	0,613	1,847	3,907	0,000
1441-1	1,554	0,356	0,127	0,856	2,253	4,360	0,000
1441-2	2,009	0,320	0,102	1,383	2,636	6,285	0,000
1441-3	0,680	0,330	0,109	0,033	1,328	2,059	0,039
1441-4	1,858	0,361	0,130	1,151	2,566	5,147	0,000
1441-5	1,129	0,299	0,090	0,542	1,716	3,771	0,000
1446-1	1,900	0,349	0,122	1,215	2,584	5,440	0,000
1446-2	1,535	0,308	0,095	0,932	2,138	4,988	0,000
1446-3	1,103	0,264	0,070	0,585	1,621	4,171	0,000
1446-4	1,745	0,331	0,110	1,096	2,394	5,269	0,000
1447-1	6,581	0,633	0,401	5,340	7,821	10,394	0,000
1447-2	5,195	0,523	0,273	4,170	6,220	9,936	0,000
1451,000	2,016	0,213	0,045	1,598	2,433	9,463	0,000
1452,000	2,613	0,264	0,070	2,095	3,131	9,880	0,000
1501,000	1,247	0,375	0,141	0,512	1,983	3,323	0,001
1502,000	1,105	0,273	0,075	0,570	1,641	4,046	0,000
1503-1	1,289	0,186	0,035	0,925	1,653	6,937	0,000
1503-2	1,068	0,181	0,033	0,714	1,422	5,909	0,000
1506,000	0,909	0,170	0,029	0,576	1,242	5,351	0,000
1507,000	3,185	0,489	0,239	2,227	4,142	6,518	0,000
1508,000	1,080	0,277	0,076	0,538	1,621	3,904	0,000
1601,000	14,929	1,387	1,924	12,211	17,648	10,763	0,000
1701,000	1,312	0,285	0,081	0,754	1,870	4,609	0,000
1703,000	2,028	0,389	0,151	1,265	2,791	5,212	0,000
1704,000	1,011	0,237	0,056	0,546	1,477	4,258	0,000
1705,000	1,162	0,374	0,140	0,430	1,894	3,110	0,002
1707,000	0,978	0,314	0,099	0,363	1,593	3,115	0,002
1902,000	1,093	0,243	0,059	0,617	1,569	4,500	0,000
1903-1	0,550	0,154	0,024	0,247	0,852	3,565	0,000
1903-2	1,024	0,161	0,026	0,708	1,339	6,362	0,000
1905,000	1,231	0,315	0,099	0,613	1,848	3,907	0,000
4001-1	0,824	0,286	0,082	0,263	1,385	2,879	0,004
4001-2	0,378	0,249	0,062	-0,109	0,866	1,521	0,128
	1,198	0,022	0,000	1,155	1,240	55,150	0,000



APPENDIX B

The forest plot for reliability coefficient method

Study name	Statistics for each study						
	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
1001,000	1,200	1,111	1,235	-0,978	3,378	1,080	0,280
1003,000	0,921	1,085	1,176	-1,205	3,047	0,849	0,396
1004,000	1,127	1,111	1,235	-1,051	3,304	1,014	0,311
1005,000	1,312	1,104	1,220	-0,853	3,476	1,188	0,235
1006,000	2,878	1,037	1,075	0,846	4,911	2,776	0,006
1007,000	0,907	1,179	1,389	-1,403	3,217	0,770	0,442
1008,000	0,064	1,043	1,087	-1,980	2,107	0,061	0,951
1009,000	1,109	1,140	1,299	-1,125	3,342	0,973	0,331
1010,000	2,326	1,147	1,316	0,078	4,574	2,028	0,043
1012,000	1,119	1,162	1,351	-1,159	3,397	0,963	0,336
1015-1	1,629	1,043	1,087	-0,415	3,672	1,562	0,118
1015-2	2,179	1,066	1,136	0,090	4,269	2,044	0,041
1016,000	3,712	1,125	1,266	1,507	5,917	3,299	0,001
1017,000	1,306	1,280	1,639	-1,203	3,816	1,020	0,308
1020,000	0,797	1,054	1,111	-1,269	2,863	0,756	0,450
1022,000	1,431	1,132	1,282	-0,788	3,650	1,264	0,206
1023,000	1,578	1,179	1,389	-0,732	3,887	1,339	0,181
1024,000	0,803	1,132	1,282	-1,417	3,022	0,709	0,478
1026-1	1,821	1,076	1,159	-0,289	3,931	1,692	0,091
1026-2	1,105	1,165	1,357	-1,178	3,388	0,949	0,343
1028,000	2,033	1,155	1,333	-0,230	4,296	1,761	0,078
1031,000	1,860	1,226	1,504	-0,544	4,263	1,517	0,129
1034,000	1,240	1,124	1,263	-0,963	3,442	1,103	0,270
1035,000	3,862	1,048	1,098	1,809	5,915	3,686	0,000
1036,000	2,315	1,155	1,333	0,052	4,578	2,005	0,045
1039,000	1,101	1,132	1,282	-1,118	3,321	0,973	0,331
1040,000	1,650	1,098	1,205	-0,501	3,802	1,504	0,133
1041-1	1,058	1,043	1,087	-0,985	3,101	1,015	0,310
1041-2	0,613	1,195	1,429	-1,730	2,956	0,513	0,608
1041-3	1,075	1,078	1,163	-1,038	3,189	0,997	0,319
1041-4	0,910	1,104	1,220	-1,254	3,075	0,824	0,410
1042-1	1,517	1,143	1,307	-0,724	3,758	1,327	0,184
1042-2	2,022	1,143	1,307	-0,219	4,263	1,769	0,077
1042-3	2,016	1,143	1,307	-0,225	4,257	1,763	0,078
1043,000	1,940	1,336	1,786	-0,679	4,559	1,452	0,147
1048,000	1,369	1,179	1,389	-0,940	3,679	1,162	0,245
1049,000	2,573	1,065	1,134	0,486	4,660	2,417	0,016
1051,000	0,813	1,072	1,149	-1,288	2,914	0,758	0,448
1052-1	1,051	1,054	1,111	-1,015	3,117	0,997	0,319
1052-2	0,518	1,054	1,111	-1,548	2,584	0,491	0,623
1052-3	1,331	1,054	1,111	-0,735	3,397	1,263	0,207
1054,000	0,588	1,179	1,389	-1,721	2,898	0,499	0,618
1055,000	0,818	1,104	1,220	-1,346	2,983	0,741	0,459
1056,000	0,916	1,140	1,299	-1,317	3,150	0,804	0,421
1059,000	1,980	1,054	1,111	-0,086	4,046	1,878	0,060
1061,000	1,478	1,132	1,282	-0,741	3,697	1,305	0,192
1065-1	0,997	1,098	1,205	-1,154	3,149	0,909	0,364
1065-2	0,786	1,059	1,121	-1,289	2,861	0,742	0,458
1067,000	0,574	1,179	1,389	-1,735	2,884	0,487	0,626
1072-1	2,386	1,111	1,235	0,208	4,564	2,147	0,032
1072-2	1,966	1,015	1,030	-0,023	3,955	1,937	0,053
1076,000	0,117	1,147	1,316	-2,131	2,365	0,102	0,919
1077,000	0,686	1,264	1,597	-1,791	3,164	0,543	0,587
1079,000	0,560	1,222	1,493	-1,835	2,954	0,458	0,647
1080,000	0,141	1,195	1,429	-2,202	2,483	0,118	0,906
1081,000	1,192	1,078	1,163	-0,922	3,305	1,105	0,269
1082,000	0,085	1,111	1,235	-2,092	2,263	0,077	0,939
1084-1	0,767	1,140	1,299	-1,467	3,001	0,673	0,501
1084-2	0,171	1,140	1,299	-2,062	2,405	0,150	0,881
1087,000	1,111	1,280	1,639	-1,398	3,621	0,868	0,386
1088,000	1,217	1,060	1,124	-0,860	3,295	1,148	0,251
1089,000	0,728	1,072	1,149	-1,374	2,829	0,679	0,497
1092,000	2,087	1,091	1,190	-0,052	4,225	1,912	0,056
1093,000	1,296	1,145	1,311	-0,948	3,540	1,132	0,258
1094,000	2,130	1,055	1,112	0,063	4,198	2,020	0,043
1095,000	2,516	1,140	1,299	0,282	4,750	2,208	0,027
1099,000	1,406	1,170	1,370	-0,888	3,700	1,201	0,230
1102,000	1,162	1,181	1,395	-1,153	3,476	0,984	0,325
1103,000	0,728	1,240	1,538	-1,703	3,159	0,587	0,557
1104,000	1,107	1,179	1,389	-1,203	3,417	0,939	0,348
1110,000	1,200	1,098	1,205	-0,951	3,351	1,093	0,274
1113,000	0,535	1,052	1,106	-1,527	2,596	0,508	0,611
1115,000	0,875	1,159	1,344	-1,397	3,148	0,755	0,450
1116,000	0,129	1,062	1,129	-1,953	2,211	0,121	0,903
1117-1	0,138	1,132	1,282	-2,081	2,358	0,122	0,903
1117-2	0,808	1,132	1,282	-1,411	3,027	0,714	0,475
1118,000	1,271	1,060	1,124	-0,807	3,348	1,199	0,231
1126,000	2,576	1,292	1,669	0,044	5,108	1,994	0,046
1127,000	2,252	1,071	1,147	0,154	4,351	2,103	0,035
1903-E	1,650	1,118	1,250	-0,541	3,842	1,476	0,140
1301,000	1,233	1,187	1,408	-1,093	3,559	1,039	0,299
1303,000	1,529	1,250	1,563	-0,921	3,979	1,224	0,221
1401,000	0,050	1,179	1,389	-2,260	2,359	0,042	0,966
1405,000	1,857	1,132	1,282	-0,362	4,076	1,640	0,101
1406,000	2,298	1,118	1,250	0,106	4,489	2,055	0,040
1409,000	1,255	1,078	1,163	-0,859	3,368	1,163	0,245
1410,000	0,969	1,157	1,339	-1,299	3,236	0,837	0,402
1412,000	0,819	1,091	1,190	-1,319	2,958	0,751	0,453
1413-1	1,627	1,125	1,266	-0,578	3,833	1,446	0,148
1413-2	2,744	1,125	1,266	0,538	4,949	2,439	0,015
1413-3	1,573	1,125	1,266	-0,632	3,778	1,398	0,162
1413-4	1,203	1,125	1,266	-1,002	3,408	1,069	0,285
1413-5	1,642	1,125	1,266	-0,564	3,847	1,459	0,145
1414,000	1,475	1,118	1,250	-0,716	3,667	1,320	0,187
1416,000	0,807	1,218	1,484	-1,581	3,194	0,662	0,508
1422,000	0,868	1,162	1,351	-1,410	3,147	0,747	0,455
1423,000	1,065	1,195	1,429	-1,278	3,407	0,891	0,373
1424,000	0,861	1,073	1,151	-1,241	2,964	0,803	0,422
1425,000	4,893	1,124	1,264	2,690	7,097	4,352	0,000
1428,000	0,575	1,085	1,176	-1,551	2,701	0,530	0,596
1432-1	1,882	1,155	1,333	-0,381	4,145	1,630	0,103
1433-1	0,070	1,072	1,149	-2,031	2,171	0,065	0,948
1433-2	0,923	1,072	1,149	-1,178	3,025	0,861	0,389
1433-3	1,459	1,072	1,149	-0,642	3,560	1,361	0,174
1435,000	0,849	1,125	1,266	-1,356	3,054	0,755	0,450
1436,000	1,478	1,111	1,235	-0,700	3,656	1,330	0,183
1437-1	0,896	1,348	1,818	-1,747	3,539	0,665	0,506
1437-2	1,034	1,348	1,818	-1,609	3,676	0,767	0,443
1437-3	0,813	1,348	1,818	-1,830	3,456	0,603	0,546
1438-1	0,823	1,104	1,220	-1,342	2,987	0,745	0,456
1438-2	0,823	1,104	1,220	-1,342	2,987	0,745	0,456
1439-1	1,215	1,336	1,786	-1,404	3,834	0,909	0,363
1439-2	1,008	1,336	1,786	-1,611	3,628	0,755	0,450
1439-3	1,376	1,336	1,786	-1,243	3,995	1,030	0,303
1439-4	0,723	1,336	1,786	-1,896	3,342	0,541	0,589
1440-1	1,215	1,125	1,266	-0,990	3,420	1,080	0,280
1440-2	1,230	1,125	1,266	-0,975	3,435	1,093	0,274
1441-1	1,554	1,069	1,143	-0,541	3,650	1,454	0,146
1441-2	2,009	1,069	1,143	-0,086	4,104	1,879	0,060
1441-3	0,680	1,036	1,074	-1,351	2,711	0,656	0,512
1441-4	1,858	1,036	1,074	-0,173	3,889	1,793	0,073
1441-5	1,129	1,036	1,074	-0,902	3,160	1,089	0,276
1446-1	1,900	1,179	1,389	-0,410	4,209	1,612	0,107
1446-2	1,535	1,147	1,316	-0,714	3,783	1,338	0,181
1446-3	1,103	1,048	1,099	-0,952	3,158	1,052	0,293
1446-4	1,745	1,098	1,205	-0,406	3,897	1,590	0,112
1447-1	6,581	1,147	1,316	4,332	8,829	5,737	0,000
1447-2	5,195	1,140	1,299	2,961	7,429	4,559	0,000
1451,000	2,016	1,222	1,493	-0,379	4,410	1,650	0,099
1452,000	2,613	1,153	1,330	0,353	4,873	2,266	0,023
1501,000	1,247	1,170	1,370	-1,047	3,541	1,066	0,287
1502,000	1,105	1,165	1,357	-1,178	3,388	0,949	0,343
1503-1	1,289	1,043	1,087	-0,755	3,332	1,236	0,216
1503-2	1,068	1,125	1,266	-1,137	3,273	0,949	0,343
1506,000	0,909	1,195	1,429	-1,434	3,251	0,760	0,447
1507,000	3,185	1,098	1,205	1,033	5,336	2,901	0,004
1508,000	1,080	1,213	1,471	-1,297	3,456	0,890	0,373
1601,000	14,929	1,121	1,256	12,733	17,126	13,320	0,000
1701,000	1,312	1,147	1,316	-0,936	3,560	1,144	0,253
1703,000	2,028	1,132	1,282	-0,191	4,247	1,791	0,073
1704,000	1,011	1,170	1,370	-1,283	3,305	0,864	0,388
1705,000	1,162	1,091	1,190	-0,977	3,300	1,065	0,287
1707,000	0,978	1,132	1,282	-1,241	3,197	0,864	0,388
1902,000	1,093	1,110	1,232	-1,082	3,268	0,985	0,325
1903-1	0,550	1,179	1,389	-1,760	2,859	0,466	0,641
1903-2	1,024	1,162	1,351	-1,255	3,302	0,881	0,379
1905,000	1,231	1,111	1,235	-0,947	3,409	1,108	0,268
4001-1	0,824	1,091	1,190	-1,314	2,963	0,755	0,450
4001-2	0,378	1,091	1,190	-1,760	2,517	0,347	0,729
	1,476	0,093	0,009	1,295	1,658	15,952	0,000


Std diff in means and 95% CI
Favours A
Favours B