Gender Effect on Competency Perceptions on Measurement and Evaluation: A Meta-Analysis Study

Merve ŞAHİN KÜRŞAD
Milli Savunma Üniversitesi

ABSTRACT

The purpose of this research is to examine the effect of gender on measurement and evaluation competency using meta-analysis. By statistically combining data obtained from various studies on a given subject, meta-analysis attempts to reach a general conclusion on the results of these studies. For this purpose, studies on general competency perceptions on measurement and evaluation were collected by a literature review and 17 studies were included in the analysis. Effect sizes were calculated in 95% confidence intervals for these studies. Studies were evaluated according to the random effects model. Gender was found out to be an ineffective independent variable in competency perceptions on measurement and evaluation.

Keywords: competency on measurement and evaluation, gender, meta-analysis

Gönderim: 24.09.2021
Kabul: 17.11.2021
Yayımlanma: 28.02.2022

Sorumlu Yazar:
Merve Şahin Kürşad
sahinmerv@gmail.com

Cinsiyetin Ölçme ve Değerlendirmeye Yönelik Genel Yeterlik Algısı Üzerindeki Etkisi: Bir Meta Analiz Çalışması

Merve ŞAHİN KÜRŞAD
Milli Savunma Üniversitesi

ÖZET

Bu çalışmanın amacı cinsiyetin ölçme ve değerlendirmeye yönelik genel yeterlik üzerindeki etkisini meta analiz yöntemiyle incelemektir. Meta analiz bir konuda yapılmış olan farklı birçok çalışmadan elde edilen verileri istatistiksel olarak bir araya getirerek, bu çalışma sonuçları hakkında genel bir yargıya varma amacıyla kullanılmaktadır. Bu amaçla ölçme ve değerlendirmeye yönelik genel yeterlik algısı üzerine yapılmış olan çalışmalar, alanyazın taraması yapılarak bir araya getirilmiş ve toplamda 17 çalışma analize dâhil edilmiştir. Bu çalışmalar için %95 güven aralığında etki büyüklükleri hesaplanmıştır. Çalışmalar rastgele etkiler modeline göre değerlendirilmiştir. Sonuç olarak cinsiyetin, ölçme ve değerlendirmeye yönelik genel yeterlik algısı üzerinde baskın bir bağımsız değişken olmadığı belirlenmiştir.

Anahtar kelimeler: ölçme ve değerlendirmeye yönelik genel yeterlik, cinsiyet, meta-analiz

Received: 24.09.2021
Accepted: 17.11.2021
Published: 28.02.2022

Corresponding Author:
Merve Şahin Kürşad
sahinmerv@gmail.com

INTRODUCTION

One of the main components of education is the competence of pre-service teachers and teachers. Competency is the required attitude, knowledge and ability to fulfill duties related to an occupation in a qualified manner (MEB, 2008). Competencies affect investment, targets and the level of attaining goals in education (Tschannen-Moran & Hoy, 2001). The education model will be qualified to the extent that teachers are qualified in their occupations (Akçamete, 2005). One of the important points in training teachers is the fact that society and therefore education requirements change continuously. The qualifications teachers should possess for attaining education goals in this continuous process form teachers' competency areas (Şahin, 2004). Within these competency areas, the competencies teachers should possess for evaluating students, specifying to what extent the determined goals and behaviors were reached and pointing out deficiencies fall under measurement and evaluation competencies (Atılgan, 2007). The Ministry of National Education (MoNE) places measurement and evaluation under the header, "Teachers' General Competencies" as sub-competence, "Measurement and Evaluation".

Aside from the information from MoNE, the American Foundation of Teachers, National Council on Measurement in Education and National Education Association (1990) make the following comments on teachers' competencies for measurement and evaluation: Teachers should possess the skill to choose measurement and evaluation methods appropriate for their classes, skill to develop measurement and evaluation methods, skill to score, apply and interpret the developed measurement and evaluation methods, skill to make use of the measurement and evaluation results when deciding on teaching planning, student and curriculum development, skill to develop methods for student grading applicable in measurement and evaluation. Teachers should be qualified to notify students, parents, other concerned parties and educators of the measurement and evaluation results.

A well-qualified teacher in her/his field should know the appropriate measurement and evaluation processes and tools to use when assessing students regarding whether they have reached targeted goals. In addition, a qualified teacher should know how to use the measurement and evaluation tools in line with students' learning gaps, intended for eliminating them (Karaca, 2003). In fact, measurement and evaluation aims at bettering the quality of the education process (Balcı & Tekkaya, 2000). As teachers' qualifications determine education quality, the system's success directly depends on the competencies of the teacher responsible (Genç, 2008). Therefore, these competencies should be acquired by pre-service teachers through theoretical and practical studies (Karaca, 2003).

A review of the literature shows that measurement and evaluation competency is related to various factors including department and class level (Günal, Usta & Uluman, 2015; Pektaş, 2010). In addition, the gender variable was observed to be another significant variable related to measurement and evaluation competency and which is frequently examined by researchers (Akdağ, 2011; Altun, 2017; Eğri, 2006; Erdoğdu & Kurt, 2012; Günal et. al.2015; Pektaş, 2010; Şimşek, 2018; Yaralı, 2017)

In their study, examining the measurement and evaluation competencies of pre-service teachers by various variables, Günal et. al (2015) compared data by gender. The study consisting of 590 samples found that women have greater measurement and evaluation competency than men. A study by Şimşek (2018), which consisted of 685 pre-service teachers, turned up similar findings. On the other hand, in their study examining the measurement and evaluation competencies of 627 pre-service teachers, Pektaş (2010) found that men find their measurement and evaluation competency to be more adequate than women in the various sub-dimensions as well as measurement as a whole.

The study by Yaralı (2017) examined the general competency perceptions on measurement and evaluation of 413 pre-service teachers and found no statistically significant difference between women and men. The study by Altun (2017) did not find any gender-related differences when examining attitude and competency perceptions on measurement and evaluation of pre-service teachers. Studies by Akdağ (2011), Eğri (2006) and Erdoğdu and Kurt (2012) also encountered similar findings.

In the relevant literature, there are various research studies that reached different results in terms of gender. Consequently, it is not clearly known whether measurement and evaluation competency differs according to gender. Determining whether measurement and evaluation competency differs according to gender is crucial for the Ministry of National Education to take necessary precaution in case one group

has lower competency than the other. For instance, this could be taken into account when implementing activities or training sessions appropriate for gender that will improve measurement and evaluation competency. That is, different activities can be applied according to gender. In addition, determining to what extent women and men are able to perceive themselves as competent in measurement and evaluation will be a contribution to the field that will allow them to take the necessary precautions by reviewing their education process.

No consensus has been reached through comparing whether women or men have higher competencies of measurement and evaluation. Comparisons made particularly following the 2000s show that no clear result could be attained regarding the measurement and evaluation competencies of women and men. While some studies found that women have higher competency perceptions of measurement and evaluation than men (Günal et al., 2015; Şimşek, 2018), some reached the conclusion that men have higher measurement and evaluation competencies than women (Pektaş, 2010). Apart from these findings, some studies indicate that there is no difference between women and men in terms of measurement and evaluation competencies (Akdağ, 2011; Altun, 2017; Eğri, 2006; Erdoğdu & Kurt, 2012; Yaralı, 2017). The relationship between measurement and evaluation competency and gender should become clearer by systematically reviewing previous studies on the subject in an effort to eliminate the differences. As distinct from above mentioned studies, this study aims at determining whether women or men have higher measurement and evaluation competencies by systematically synthesizing the relationship between measurement and evaluation competency and gender and making use of the large size of generated samples. There is no meta-analysis study in the literature, which indicates the relationship between measurement and evaluation competency and gender, a significant societal factor comprising of women and men. In addition to the relationship between gender and measurement and evaluation competency, the type of publication variable, which might have an effect on this relationship, was included in the current study as a moderator variable. The relationship between these two variables was thus examined by meta-analysis and answers to the following research questions were sought:

1. Does gender make a meaningful difference on general competencies in terms of measurement and evaluation?

2. Does gender's effect on general measurement and evaluation competencies differ according to the type of publication (articles, theses)?

METHOD

Research Model

This study is a meta-analysis research. Referred to as the analysis of analyses, meta-analysis is used to combine multiple research results derived from individual studies (Glass, 1976). Differences in sampling design and experimental design or the existence of different variables used in studies generate results which are different from previous studies on the same subject (Hartung, Knapp & Sinha, 2008). It is rather time-consuming for a reader to reach needed information and to follow-up all studies about the subject. The aim of meta-analysis is therefore to regulate and synthesize pieces of research and obtain meaningful results from complicated and contradictory findings (Açıkel, 2009).

Data Collection

The studies included in this meta-analysis are studies carried out in Turkey between 2000 and 2020 regarding measurement and evaluation competency. The studies were chosen from among theses and articles. Review of these studies was carried out both in Turkish and in English. Review of studies were searched via Google Scholar search engine and the Council of Higher Education using terms, “ölçme ve değerlendirmeye yönelik yeterlik”, “ölçme ve değerlendirme yeterlik”, “ölçme-değerlendirmeye yönelik yeterlik", "ölçme ve değerlendirmeye yönelik yeterlik algısı", "ölçme ve değerlendirme",

"yeterlik", "competency perceptions on measurement and assessment", "competency perceptions on measurement and evaluation" and "competency on measurement and assessment". Searches yielded 47 studies. Two of the studies could not be accessed, 25 of them did not possess gender-related findings or they did not report the total score of the scale by gender, only its sub-dimensions. 20 studies were therefore included in the analysis and 3 of them were eliminated during analysis related with normality factor as mentioned in the findings section. Criteria for inclusion in the analysis are as follows:

1. The study should have been carried out between 2000 and 2020.
2. The study should be in a thesis or article format published in Turkish or English.
3. The study should make a comparison of measurement and evaluation competency in terms of gender and these comparisons should be made using a total score of the scale.
4. The study should report on the size of the sample group, standard deviation and mean value.

Coding of Data and Ensuring Validity and Reliability

A coding form was developed by the researcher to determine whether the studies are to be included in the meta-analysis. Information in the form was determined with the aim of enlisting the characteristics of the studies. The coding form contains information regarding the name of study, author of study, type of study, year of publication, standard deviation of competency perceptions on measurement and evaluation, mean value of competency perceptions on measurement and evaluation and size of the sample group.

Two lecturers from Turkish Language and Measurement and Evaluation departments were provided with information about the aim and scope of the study to ensure content validity of the coding form and the form was finalized in accordance with these experts' opinion.

Coding was carried out independently by two researchers for the sake of reliability of the study. The number of mutual coding was determined by comparing researchers' analysis and reliability level was calculated with the reliability level formula developed by Miles and Huberman (2002). Reliability level of the coding was found to be 98 per cent. Reliability levels of 70 per cent and above obtained from this formula is stated to be reliable. The coding was therefore reliable.

Data Analysis and Interpretation

This study used group differences meta-analysis. In group differences meta-analysis, standardized effect size values are calculated and comparisons are made to determine differences among groups (Durlak, 1995). Hedges' g effect size was used to test differences in terms of gender in the studies, which were included in the meta-analysis, due to the use of different scales and the possibility of obtaining an extreme value. This effect size value is used to observe differences between groups via standard deviations adjusted for sample size (Ware, Kirkovski & Lum, 2020). The formula for the Hedges' g coefficient used in this study is as follows:

$$\text{Hedges' g} = \frac{M1 - M2}{\text{Spooled}}$$

$$S_{pooled} = \sqrt{\frac{(n1-1)S12 + (n2-1)S22}{n1+n2-2}}$$

The Hedges' g formula indicates the difference between means and indicates pooled standard deviation. Data used in independent studies are converted into a mutual unit of measurement using this formula (Borenstein, Hedges, Higgins & Rothstein, 2009; Eells, 2011). Since studies included in this meta-

analysis used different scales and samples, effect sizes obtained from each study were calculated separately. The interpretation of obtained effect sizes according to Cohen (1988, p. 82) is as follows: If the effect size is 0.20 and below, the effect is small; if it is between 0.20 and 0.80, the effect is medium and if it is 0.80 and above, the effect is large. Comprehensive Meta Analysis (CMA) program, which was developed by Borenstein et al. (2009), was used in calculating effect sizes and forest graph is drawn using the distribution of effect sizes. This graph indicates weights of studies in the meta-analysis (Kılıç, 2016).

A literature review was conducted to determine the studies to be included in the meta-analysis during the analysis of the data. The study sample only contains theses and articles. Before calculating effect sizes in the meta-analysis, Q statistic, which was developed by Hedges and Olkin (1985) and is used to test the heterogeneity of samples and effect sizes, was calculated. It was then decided whether the analysis was to be evaluated according to the random effects model or the fixed effects model. Q statistic uses chi-square distribution and heterogeneity is decided by comparison with the critical chi-square values (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Aside from the Q statistic, F values, another indicator for heterogeneity, were calculated. Heterogeneity analysis is a low power analysis. It means that even if heterogeneity is available in certain circumstances, it might not be identified. Thus, the value is reported along with the Q value and it indicates the measure of heterogeneity (Crombie & Davies, 2009).

Following the evaluation of the heterogeneity-related indicators, funnel plot, Begg and Mazumdar rank correlation test and Orwin's fail-safe N value were used to evaluate publication bias. The most comprehensive way to identify publication bias is by creating a funnel plot. Funnel plots are a kind of scatter plot. They have effect sizes on the horizontal axis and sample sizes, variables or standard error of the studies on the vertical axis. Funnel plots, in which values are concentrated around mean effect sizes and which display a symmetrical distribution, could be said to be free of publication bias (Copas & Shi, 2000).

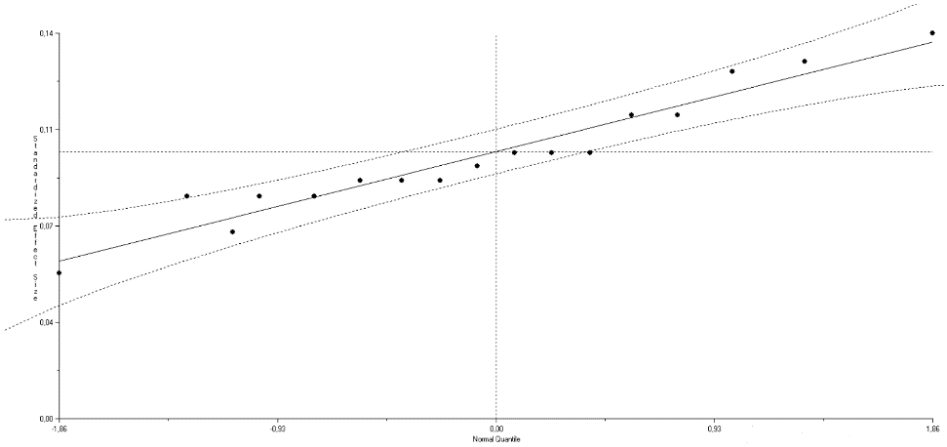
In certain circumstances, funnel plots might not identify publication bias. In fact, funnel plots display a visual graphic about publication bias. Another alternative for examining bias is to calculate the fail-safe number. This number estimates the number of studies, which would cause the results obtained from the meta-analysis to change. In other words, it estimates the number of studies, which would cause the general tendency of the studies used in the meta-analysis to change. This number is significant for evaluating the reliability of the meta-analysis results (Lipsey & Wilson, 2001). Fail-safe N values of Rosenthal or Orwin are accordingly evaluated. Rosenthal's fail-safe N calculates how many studies should be added to the analysis in order to reduce the p-value to a trivial size. In other words, it identifies how many studies are available which would increase the significance level of the effect above 0.05 or 0.01 (Borenstein et al., 2009).

Aside from fail-safe N value, another indicator of publication bias is Begg and Mazumdar rank correlation test. This test is an indicator proving the study sample was not chosen with bias (Dağyar & Demirel, 2015). Kendall's tau significance level being $p>0.05$ thus indicates that there is no publication bias. The results of this test were reported accordingly.

MetaWin and Comprehensive Meta Analysis (CMA) 2.0 programs were used when analyzing data. MetaWin program was used to evaluate normality of studies included in the meta-analysis. CMA 2.0 program was used to calculate general effect size, forest graph, funnel plots and publication bias calculations.

FINDINGS

In this section, firstly normality and then meta-analysis results will be reported. Before conducting meta-analysis, the compatibility of the effect sizes of the study with normal distribution should be examined. This assumption was checked using the MetaWin program. Graph 1 below shows outputs about the normality assumptions.



Graph 1. *Effect sizes distribution of studies according to the gender variable*

When normality outputs in Graph 1 are examined, the general distribution is observed to fall along the X=Y line and within the boundary lines displayed as confidence intervals. Two out of 20 studies examined in accordance with the gender variable were eliminated from analysis due to skewing the normality and one was eliminated due to being an extreme value. Meta-analysis was carried out for 17 studies. The examined distribution of studies shows that they fall within the boundary lines displayed as confidence intervals.

Table 1 shows findings of general effect sizes and the homogeneity test regarding measurement and evaluation competency by the gender variable.

Table 1. *Findings of gender variable according to effect sizes of fixed and random effects models*

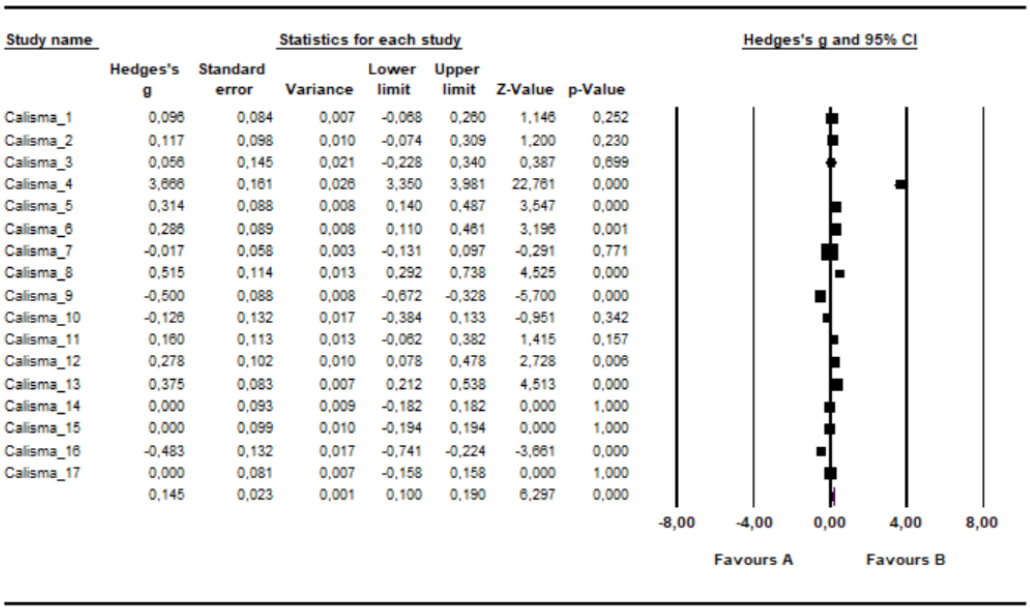
								95 per cent Confidence Interval for Effect Sizes	
	Degrees of Freedom	z	p	Standard Error	Hedges' *g*	Q	I^2	Bottom	Top
Fixed Effects Model	16	6.296	0.000	0.023	0.145	601.323	97.339	0.100	0.190
Random Effects Model	16	1.901	0.057	0.143	0.271			-0.008	0.551

When outputs of the homogeneity test in Table 1 are examined, the Q value is as 601.323. Since this value is larger than the critical chi-squared value of 26.296, equivalent to 16 degrees of freedom in 95 per cent confidence interval, it can be stated that the distribution of studies is heterogeneous in terms of gender. The value indicating the size of heterogeneity (I^2) was calculated as 97.339 per cent. For moderate, substantial and considerable levels of heterogeneity, 25 per cent, 50 per cent and 75 per cent limit values are suggested respectively (Higgins, Thompson, Deeks & Altman, 2003). It can thus be said that this study has a considerable level of heterogeneity.

Due to the studies turning out to be heterogeneous, errors originating from the sample being heterogeneous can be eliminated using the random effects model (Ayaz & Söylemez, 2015). The mean effect size value was found to be 0.271 with 0.143 margin of error using the random effects model. The lower limit of the effect size was found out to be -0.008 and the upper limit was found out to be 0.551 in the 95 per cent confidence interval. The random effects model did not reveal statistically meaningful

differences in the distribution of effect sizes between studies in terms of gender (p> 0.05). Therefore, studies included in this meta-analysis show no differences between women and men for measurement and evaluation competencies in terms of gender.

Graph 2 shows the forest graph indicating the effect size distributions according to the random effects model.



Graph 2. *Forest graph of effect size distributions of the studies*

Squares in Graph 2 indicate effect sizes of the studies and the size of the squares indicates the weight of the studies in the general effect sizes. It can thus be seen that the study with the most weight is "Calisma_7". When effect sizes of the studies were examined, it was found that four of 17 studies have negative and 13 have positive effects. As can be seen from the distribution of the studies, no tendency was observed towards women or men.

Type of publication was also included as the moderator variable in the analysis examining differences in terms of gender. Table 2 shows findings on whether measurement and evaluation competencies differ according to the type of publication.

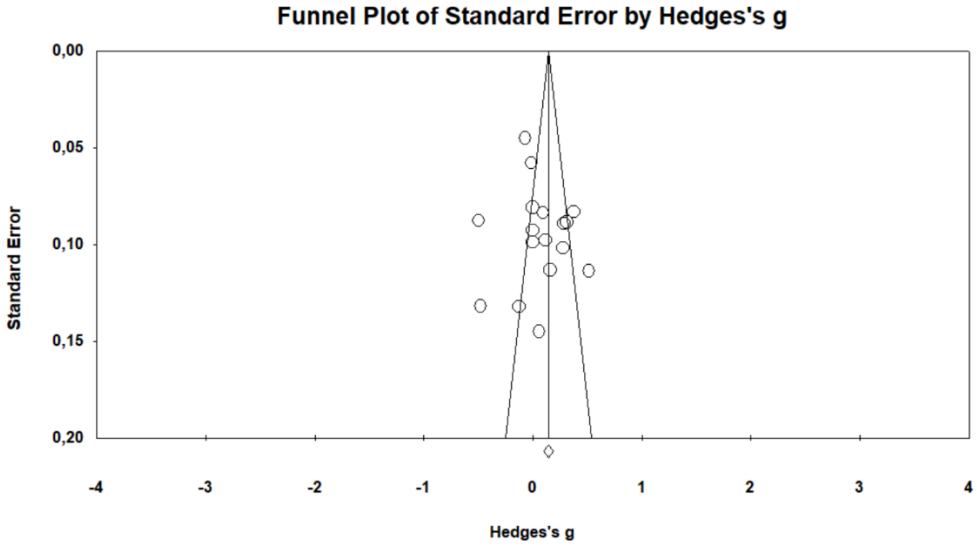
Table 2. *Effect size values according to types of publication*

Variable	Homogeneity Value Among Groups	p	Degrees of Freedom	N	Hedges' *g*	Standard Error	95 per cent Confidence Interval for Effect Sizes	
Type of Publication	3.065	0.310	1	17			Lower Limit	Upper Limit
Thesis				9	0.512	0.214	-0.420	0.417
Article				8	-0.001	0.201	0.118	0.906

Chi-squared value equivalent to one degree of freedom in 95 per cent confidence interval is 3.841. This value is larger than the Q value in Table 2. In other words, the homogeneity value among groups is

smaller than the critical value. The groups formed according to types of publication thus have homogenous distribution. The examined significance value of effect sizes among the groups is found to be 0.310 and p> 0.05. This means that there are no meaningful differences between the distributions of effect sizes. Therefore, results obtained for measurement and evaluation competencies in aspect of gender variable do not change according to the type of publication. In other words, it can be stated that variance is not caused by the type of publication.

Graph 3 shows the funnel plot for publication bias of studies included in the meta-analysis in terms of the gender variable.



Graph 3. *Funnel plot for publication bias of studies included in the analysis in terms of gender*

Graph 3 on publication bias displays a distribution in the upper parts of the funnel plot. In case of publication bias, the studies should gather in the lower parts of the graph in one direction (Copas & Shi, 2000). Studies included in this meta-analysis are therefore not subject to publication bias. To obtain statistically clearer information on publication bias, the results of Begg and Mazumdar rank correlation test were also examined. Table 3 has related results.

Table 3. *Biasness in studies included in the meta-analysis*

Variable	Biasness	
Gender	Kendall's S (P-Q)	20.000
	Kendall's Tau	0.147
	Z-value for Tau	0.823
	p	0.410

Table 3 shows that the studies included in the meta-analysis in terms of the gender variable are not biased (tau=0.147, p> 0.05).

Similarly, Rosenthal's fail-safe N value was used to determine publication bias. Table 4 presents the results.

Table 4. Results of Rosenthal's fail-safe N value test indicating biasness in studies of the meta-analysis

Z-value of Studies	8.444
P-value of Studies	0.000
Alpha	0.050
Direction	2.000
Z-value for Alpha	1.959
Number of Studies Examined	17
Fail-Safe Number	299

p<.05

The p-value being smaller than the alpha-value according to Rosenthal's fail-safe N test in Table 4 shows that the study is strong and reliable (Rosenthal, 1979). According to the findings, to invalidate the results of the meta-analysis, 299 more individual studies should be included in the analysis (p<.05).

CONCLUSION AND DISCUSSION

The literature shows that the effect of gender on competency perceptions on measurement and evaluation is unclear and differing results have been obtained on this subject. In this meta-analysis, carried out to make the connection between the measurement and evaluation competencies and gender clearer by synthesizing different obtained results, it was determined that measurement and evaluation competency does not differ according to gender. In other words, neither women nor men have higher or lower measurement and evaluation competencies. In addition, the relationship between measurement and evaluation competencies and gender were examined in accordance with the type of publication and it was determined that the results do not change according to whether the publication is a thesis or an article.

Gender constitutes the independent variable of many studies carried out in the social sciences. Despite criticism of gender becoming a subject to the studies, it is stated that certain behaviors can change in time with various factors, including age. Education, technology and social transformation can cause different gender groups to exhibit different behaviors. This situation requires reviewing study findings according to their sample group (Tuncer & Dikmen, 2016). Opinions and values about gender can differ across societies and time. Thus, when evaluating any situation in a study, the effect of the gender variable also comes under examination (Dikmen & Çağlar, 2015). In our context, as studies evaluated the gender variable for different groups, different context at different times, divergent results may have been obtained. Although some studies revealed differences in favor of women or men (Günal et al., 2015; Pektaş, 2010; Şimşek, 2018), these differences are no longer meaningful when the studies were combined. Since the majority of studies carried out within the literature revealed no meaningful differences (Akdağ, 2011; Altun, 2017; Eğri, 2006; Erdoğdu & Kurt, 2012; Yaralı, 2017), the results of the meta-analysis obtained through the combined analysis of these studies yielded no meaningful differences.

A process of measurement and evaluation is only successful if the teacher and students are involved and if it can generate feedback in the education process (Black, Harrison, Lee, Marshall & William, 2003, p. 2). At this point, the competency level of the teacher can be said to be directly proportional with the quality of her/his teaching process and the teacher thus should be equipped with skills on methods and techniques of measurement and evaluation (Çakan, 2004). As a matter of fact, measurement and evaluation aims at measuring the quality of education and the betterment of it (Balcı & Tekkaya, 2000). This study shows that there are no meaningful differences between women and men in this process.

Although this study presents the results synthesized from various other studies, it still cannot explain why the gender variable is frequently resorted to when examining measurement and evaluation competencies. Researchers might have used their findings of the gender variable in a supplementary training session to determine whether an arrangement of theoretical and practical activities is needed in terms of gender (Tuncer & Dikmen, 2016). However, for whatever reason, it was observed that gender was not an effective independent variable on competency perceptions on measurement and evaluation. It is recommended that further studies focus on other independent variables which can affect measurement and evaluation competencies. Also, only studies conducted in Turkey were included in this meta-analysis. Different cultures may have different results about this issue. Therefore, the effect of gender on measurement and evaluation can be examined by including studies in different cultures.

REFERENCES

Açıkel, C. (2009). Meta analiz ve kanıta dayalı tıptaki yeri. *Klinik Psikofarmoloji Bülteni, 19*(2), 164-172.

Akçamete, A. G. (2005). Açılış konuşması: Öğretmen yetiştirmede kalite sorunlar çalıştayı Ankara Üniversitesi Eğitim Bilimleri Fakültesi. Ankara: Tekışık Eğitim Araştırma Geliştirme Vakfı.

Akdağ, G. (2011). *Fen ve teknoloji öğretmenlerinin ölçme-değerlendirmeye ilişkin yeterlik algıları ve görüşleri* (Unpublished Master's Thesis). Institute of Science, Adıyaman University.

Altun, A. (2017). Öğretmen adaylarının ölçme ve değerlendirme dersine yönelik tutum ve yeterlik algılarının incelenmesi. *Kafkas Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 20*, 361-375

American Foundation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students. Educational Measurement: *Issues and Practice, 9*(4), 30-32.

Atılgan, H. (2007). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık.

Ayaz, M. F., & Söylemez, M. (2015). The effect of the project-based learning approach on the academic achievements of the students in science classes in Turkey: A meta-analysis study. *Education and Science, 40*(178), 255-283.

Balcı, E. & Tekkaya, C. (2000). Ölçme ve değerlendirme tekniklerine yönelik bir ölçeğin geliştirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 18*, 42–50.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Open University Press

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. United Kingdom: John Wiley and Sons, Ltd. Publication.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Copas, J., & Shi, J.Q. (2000). Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics, 1*(3), pp. 247–262

Crombie, I.K., & Davies, H.T.O. (2009). *What is meta-analysis?* London: Hayward Medical Communications.

Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: ilk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 37*(2),99-114

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

Dikmen, M., & Çağlar, A. (2015). Bilgisayar ve öğretim teknolojileri öğretmen adaylarının bilgisayar öz-yeterlik algılarının çeşitli değişkenler açısından incelenmesi. *İstanbul Ticaret Üniversitesi Sosyal Bilimleri Dergisi, 1*(28), 231-249.

Durlak, J. A. (1995). Understanding meta-analysis. In L. G. Grimm, & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 319–352). American Psychological Association.

Eğri, G. (2006). *Coğrafya öğretmenlerinin ölçme değerlendirme yapabilme yeterliliği* (Unpublished Master's Thesis). Institute of Educational Sciences, Gazi University.

Eells, R. J. (2011). *Meta-analysis of the relationship between collective teacher efficacy and student achivement* (Unpublished Doctoral Dissertation). Layola University.

Erdoğdu, M. Y., & Kurt, F. (2012). Öğretmenlerin ölçme ve değerlendirme yeterlilik algılarının bazı değişkenler açısından incelenmesi. *Elektronik Eğitim Bilimleri Dergisi, 1*(2), 23-36.

Genç, N. (2008). *Beden eğitimi öğretmenlerinin ölçme değerlendirme uygulamaları ve yeterlik algıları* (Unpublished Master's Thesis). Institute of Health Sciences, Mersin University.

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

Günal, Y., Usta, G., & Uluman, M. (2015). An investigation of attitudes of candidate teachers towards measurement and evaluation lesson against certain variables. *Procedia- Social and behavioral Sciences, 177*, 209-212

Hartung, J., Knapp, G., & Sinha,B.K. (2008). *Statistical meta-analysis with applications*. NJ: Wiley.

Hedges, L.V., & Olkin, I. (1985). *Statistical method for meta-analysis*. United Kingdom: Academic Press

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

Huedo-Medina, T., Sanchez-Meca, J., Marin-Martinez-, F., & Botella, J. (2006). Assessing heterogeneity in meta analysis: Q statistic or I2 index. *CHIP Documents, 19*.

Karaca, E. (2003). *Öğretmen adaylarının ölçme değerlendirme yeterliklerine ilişkin algıları* (Unpublished Doctoral Dissertation). Institute of Educational Sciences, Ankara University.

Kaya Uyanık, G., & Çalışkan, H. (2015). Sosyal bilgilerde ölçme ve değerlendirme. In C. Dönmez, & K. Yazıcı. (Ed). *Sosyal bilgiler öğretimi*. Ankara: Pegem Akademi.

Kılıç, S. (2016). Meta analizi anlama ve yorumlama. *Journal of Mood Disorders, 6*(2), 93-98

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Sage Publications, Inc.

MEB (2008). *Öğretmen yeterlikleri: öğretmenlik mesleği genel ve özel alan yeterlikleri*. Devlet Kitapları Müdürlüğü, Ankara.

Miles, M. B., & Huberman, A. M. (2002). *The qualitative researcher's companion*. California: Sage Publications.

Pektaş, S. (2010). *Öğretmen adaylarının ölçme ve değerlendirme yeterlik algılarının incelenmesi* (Unpublished Master's Thesis). Institute of Social Sciences, Bolu Abant Izzet Baysal University.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-664.

Şahin, A.E. (2004). Öğretmen yeterliklerin belirlenmesi. *Bilim ve Aklın Aydınlığında Eğitim Dergisi, 5*(58).

Şimşek, E. (2018). Öğretmen adaylarının ölçme ve değerlendirme genel yeterlik algıları: Atatürk Üniversitesi ilahiyat fakültesi örneği. *Amasya Üniversitesi İlahiyat Fakültesi Dergisi*, 59-80

Tschannen-Moran, M., & Hoy, A.W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783-805.

Turgut, M.F. (1988). *Eğitimde ölçme ve değerlendirme metotları* (6. Baskı). Ankara: Saydam Yayıncılık.

Tuncer, M., & Dikmen, M. (2016). Cinsiyetin tekno-pedagojik alan bilgisi üzerindeki etkisinin meta analiz yöntemiyle araştırılması. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, *28*(1), 85-92.

Ware, A.T., Kirkocski, M. & Lum, J.A.G. (2020). Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Front. Psychol., 11*(1458). Doi: 10.3389/fpsyg.2020.01458

Yaralı, D. (2017). Öğretmen adaylarının ölçme ve değerlendirmeye yönelik yeterlik algılarının incelenmesi (Kafkas Üniversitesi Örneği). *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *17*(1), 487-504.