Eurasian Journal of Educational Research 70 (2017) 37-56

Eurasian Journal of Educational Research
www.ejer.com.tr

A Meta-analysis of the Effectiveness of Alternative Assessment Techniques*

Eda GOZUYESIL[1] Isil TANRISEVEN[2]

ARTICLE INFO

Article History:

Received: 15 November 2016

Received in revised form: 16 March 2017

Accepted: 24 June 2017

DOI: http://dx.doi.org/10.14689/ejer.2017.70.3

Keywords

authentic assessment, portfolio performance, effect size

ABSTRACT

Purpose: Recent trends have encouraged the use of alternative assessment tools in class in line with the recommendations made by the updated curricula. It is of great importance to understand how alternative assessment affects students' academic outcomes and which techniques are most effective in which contexts. This study aims to examine the impact of alternative assessment techniques on achievement. **Research Methods:** In the study, a meta-analysis was conducted to combine the effect sizes of the primary studies during data collection and data analysis. **Findings:** Data analysis indicated that alternative assessment techniques have a significant and positive effect (d=0.84) on students' academic achievement. Such techniques have been found to be more effective in Mathematics courses (d=0.84), and the effect of using portfolios in class (d=1.01) is worthy of note. In accordance with the moderator analysis, whereas the effect sizes do not significantly vary in terms of subject matter and type of alternative assessment technique, there is a significant difference in the effect sizes in terms of school levels of students.

Implications for Research and Practice: The results highlighted portfolios as a highly effective assessment technique for students' academic achievements, and it revealed the impact of alternative assessment techniques on enhancing academic outcome. However, the low effectiveness of authentic assessment at the primary level may be associated with the development of creativity and critical thinking skills over time.

*This study was presented at the 3rd International Eurasian Educational Research Congress in Muğla, 31 May – 03 June, 2016.

[1] Corresponding author: Nigde Omer Halisdemir University, TURKEY, edagozuyesil@gmail.com,
ORCID: http://orcid.org/0000-0002-2094-7600.

[2] Mersin University, TURKEY, isiltanriseven77@gmail.com,
ORCID: http://orcid.org/0000-0001-5884-2807.

Introduction

A system based on the constructivist approach has been introduced with the education reform implemented in primary and secondary curricula in Turkey since 2004. The newly-developed curricula based on this approach have broken new ground in course content, teaching methods, materials and measurement and evaluation techniques (Gelbal and Kelecioglu, 2007; Yesilyurt, 2012). It appears that the most important innovation in these programs, which emphasize the individual differences between the learners, is in the field of evaluation (Coruhlu, Nas and Cepni, 2009; Yaman, 2011). With this change in the curricula, the use of performance-based alternative assessment tools as well as traditional assessment techniques has been suggested (Duban and Kucukyilmaz, 2008; Ozdemir, 2010). In this way, it has become important to evaluate students' skills and success from all aspects during the learning process and to observe their improvement.

Alternative assessment is defined as a non-traditional approach that informs students about what they know and can do, determines what they comprehend about the subject, and evaluates their performance (Gummer and Shepardson, 2001). Alternative assessment with reliable, performance-based, realistic, constructivist and feasible features includes activities in which knowledge and skill are connected and knowledge is acquired in different learning environments. It teaches students to be aware of their own ideas and to evaluate themselves by allowing students to analyze their own learning styles. In other words, alternative assessment provides flexible and meaningful learning experiences that take into consideration the learning styles of the students. From this aspect, it may be distinguished from standardized assessment techniques (Korkmaz, 2006).

Alternative assessment techniques enable students to be evaluated multi-dimensionally as they offer students multiple evaluation opportunities during which to display their knowledge, skills and attitudes (MEB, 2005). Additionally, alternative assessment assists teachers in creating a motivating learning environment that fits each student's learning needs and learning style, follows individual student achievement, and creates an atmosphere that takes into consideration students' self-assessment of their own learning process (Greenstein, 2010).

There have been many primary studies on alternative assessment techniques in Turkey. Even though the frequency of use of these techniques differs according to the subject matter (Yazici and Sozbilir, 2014), the studies have asserted that portfolios, peer assessment, diagnostic-branched trees, structured grids (Buyuktokatli and Bayraktar, 2014; Yazici and Sozbilir, 2014) and self-assessment (Karakus, 2010; Kostereliogl u and Celen, 2016) are the least-used techniques. However, the literature focuses on the positive effects of these techniques on students.

It is important that teachers use the assessment techniques recommended in their curricula to evaluate their students and their teaching activities. The current curriculum suggests that learners should be assessed in a way that will open up all-round and high-level thinking skills, and for this it provides teachers with assessment tools to evaluate students from every aspect. However, the studies (Dos, 2016; Gerek,

2006; Gelbal and Kelecioglu, 2007) show that the teachers who are unfamiliar with the curricula act with suspicion towards these techniques and view them as difficult to apply.

Many studies on the effectiveness of alternative assessment techniques have been carried out. However, no study found in either national or international literature examined the effects of alternative assessment techniques on a large scale or determined which techniques proved most effective on achievement. Accordingly, this study was designed to review the literature regarding alternative assessment that has recently gained popularity in Turkey. Data were derived from the primary studies, and the findings were combined through a meta-analysis underlying this research. Thus, calculating the effect size of the primary studies, which have investigated the impact of alternative assessment on academic outcomes, allows for the discussion of which assessment techniques are most effective.

In light of these facts, and seeing the need for this extensive review in the Turkish assessment context, the following research questions were designed for the present meta-analysis:

1. What are the effects of alternative assessment techniques on student achievement?
2. How do various alternative assessment techniques (e.g., portfolio, self-assessment) moderate the overall average effect size?
3. How do demographic features of the studies (i.e., subject matter and school level) moderate this overall effect?

Method

Research Design

The current study primarily aimed to examine the impact of alternative assessment techniques on academic achievement. In line with this purpose, a meta-analysis method was applied in this study. Meta-analysis is a statistical procedural method used to interpret, synthesize and combine the experimental findings of the primary studies on specific research (Wolf, 1986). This study was designed around Cooper's easy-to-follow seven steps for conducting a systematic review; (1) formulating the problem, (2) searching the literature, (3) gathering information from studies, (4) evaluating the quality of studies, (5) analyzing and integrating the outcomes, (6) interpreting the data, and (7) presenting the results (Cooper, 2010).

Research Instruments and Procedures

Based upon the problem of this research, extensive literature review was designed to identify the primary studies. Key words used in this review primarily consisted of "alternative assessment", "portfolio", "grid", "diagnostic tree", "peer assessment", "self-assessment" and their variations in Turkish. The following electronic databases were among the sources examined: CoHE National Dissertation Center, ERIC, PsycINFO, ASOS social sciences index and many journals of Education Faculties in

Turkey, in addition to Google web and scholar in search of conference proceedings. The primary studies were collected by regularly reviewing the databases up to August 2014, and they were selected for inclusion in the analysis. To be included, a study had to meet the following criteria:

- Address the impact of alternative assessment techniques on students' achievements,
- Contain at least two independent samples, with pretest-posttest experimental or quasi-experimental design,
- Contain sufficient statistical information to extract effect size,
- Be administered in Turkey,
- Be published between 2004 and August 2014.

As the sampling of a study must consist of at least 10 students for each group to ensure the approximate normal distribution of Cohen's d effect size (Hedges and Olkin, 1985), the studies carried out with smaller samples were not included in this analysis. In light of Lipsey and Wilson's (2001) suggestions, a coding form which included both statistical and theoretical data was developed with regard to transforming the features of all studies included in this meta-analysis into the categorical variables.

For the interrater reliability of the coding form, about 25% (n=6) of the included articles were randomly selected, and they were independently rated and coded by two researchers. The forms were compared using the [agreement / (agreement + disagreement) x 100] formula (Miles and Huberman, 1994), and the reliability of intercoders was determined to be 98%. The disagreements were discussed until they were solved and corrected on the form.

Research Sample

Subsequent to coding the studies, out of 172 theses and dissertations, 68 articles and conference papers, 26 studies (36 effect sizes) which met the criteria were identified as the sample of this meta-analytic study.

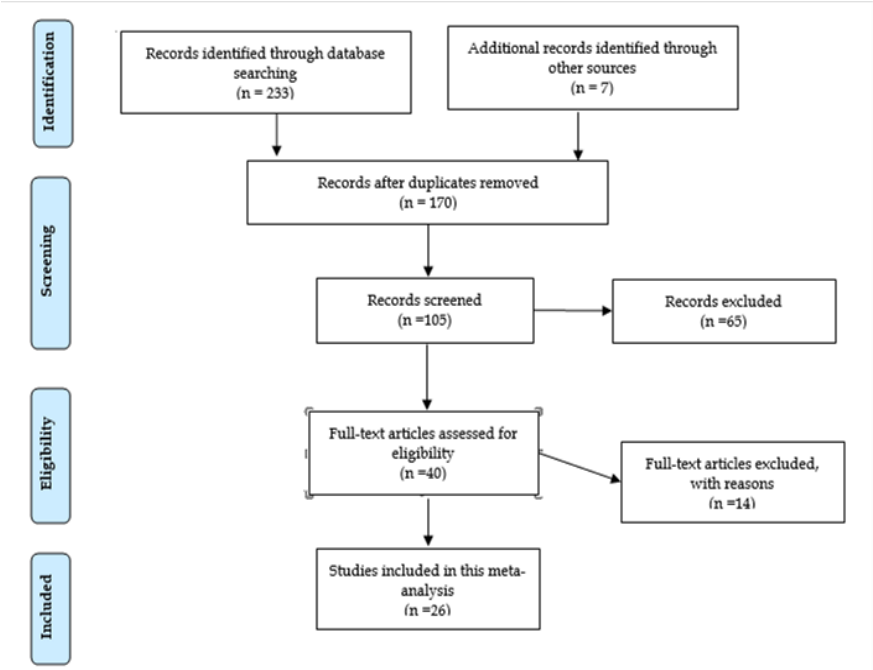

Figure 1. Flow chart of literature review

Data Analysis

The current study uses '*study effect*' meta-analysis for the analysis of the data. This method is used for group differences that occur when the arithmetical mean values of the dependent variables of each study included in meta-analysis were not obtained using the same scale (Lipsey and Wilson, 2001; Cohen, 1992). The aim of this method is to calculate the difference between the mean values of the control and experimental groups in experimental studies, represented by the formula d= (Xe-Xc)/SD (Hunter and Schmidt, 2004). The "d" value obtained represents the effect size and forms the basis for meta-analysis. In this study, the experimental group is the group to which one of the alternative assessment techniques was administered, and the control group is the one which was assessed in a traditional way. As a result, if the calculated effect size is positive, it is interpreted to be effective for alternative assessment or, if it is negative, to be effective for traditional assessment.

According to Wolf (1986), if the effect sizes of a range of independent studies are statistically significant (homogeneous), these studies may be stated to test the same hypothesis. In this case, if they are heterogeneous (statistically insignificant), it is conceivable whether each study tests the same hypothesis or not. In this paper, after extracting the effect size of each study, Q statistic suggested by Cochran was used to test the homogeneity of effect sizes. Under the fixed effect model, it was revealed that

Q value exceeded critical value. For this reason, the analysis was carried out again under the random effect model. I^2 test was also used to determine the degree of heterogeneity. Moderator variables were analyzed to explain the basis of heterogeneity. Comprehensive Meta-Analysis V2 (CMA) Software was used for all data analysis.

Results

In this paper examining the impact of alternative assessment techniques on student academic achievement, the characteristic features and effect sizes of the studies have been determined by studying the samples, standard deviations and means of 26 studies. The number of students in the studies included in the meta-analysis is 2256, 1120 of which are in the experimental groups and 1136 of which are in the controlled groups. Descriptive features of the studies included in the analysis are presented in Table 1.

Table 1

Descriptive Analysis of the Included Studies in Terms of Variables

Variables	Frequency (f)	Percentage (%)
School level		
Primary	15	57.5
Secondary	7	26.9
Undergraduate	4	15.4
Subject matter		
Science and Technology	12	46.2
Math	3	11.5
English	4	15.4
Other	7	26.9
Alternative Assessment Technique		
Self-assessment	1	2.7
Peer assessment	2	5.6
Self- and peer assessment	3	8.3
Portfolio	24	66.7
Grid	2	5.6
Diagnostic branched tree and structured grid	4	11.1

As seen in Table 1, most studies were carried out at the primary level (57.5%), and the least were at the undergraduate level (15.4%). Twelve studies (46.2%) were conducted in Science and Technology courses. Portfolios (66.7%) represented the most-used technique in the included studies. Of seven studies in which more than one assessment technique was used, four studies (11.1%) made use of diagnostic branched tree and structured grid, three studies (8.3%) used self- and peer assessment techniques together.

To find the answer to the first research question, 'What is the impact of alternative assessment techniques on students' achievement?', the studies included in this meta-analysis were integrated together with standard error and variation in the common effect size. Figure 2 shows the descriptive statistics associated with 36 effect sizes from 26 studies. The study names are presented on the left of the figure. The statistics for these 36 effect sizes, such as Hedges g, the standard error and the variance are placed in the center. On the right side of the figure, a graphic called a 'Forest plot' is presented. The effect size for each study is illustrated as a dot. The lines display the width of the confidence interval for each study. Confidence intervals spanning 0.0 on the distribution are considered to be insignificantly different from zero.

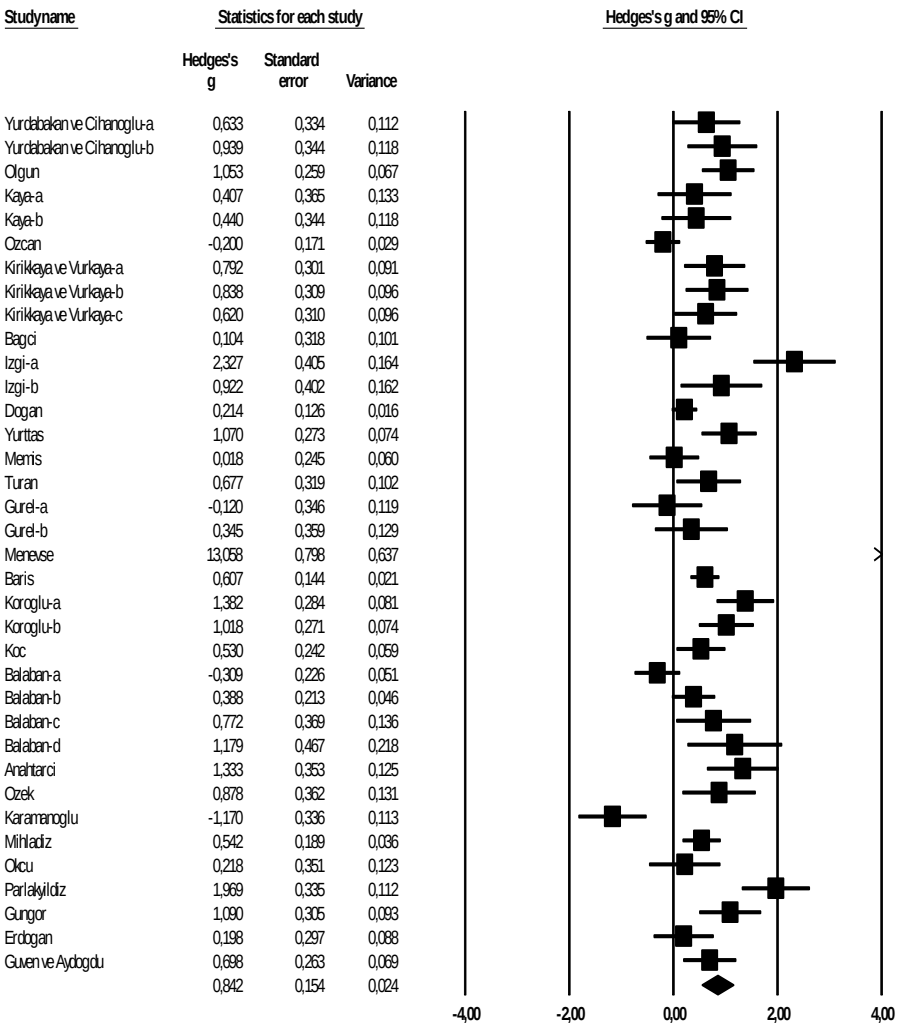

Figure 2. Forest plot of meta-analysis and study-level statistics

As reflected by the Forest plot in Figure 2, the studies with the smallest confidence interval were Koc's (2010) and Dogan's (2012), whereas the one with the widest confidence interval was Menevse's (2012). Thirty-two effect sizes from the included studies were classified as positive; that is, 88.88% of the effect sizes reveal that the results are in favor of alternative assessment techniques. The summary statistics derived from 36 effect sizes are presented in Table 2.

Table 2

Overall Weighted Average Random Effects and Fixed Effect Sizes and Homogeneity Statistics

Analytical models	N	Effect size	df	Q_{-total}	I^2	95% Confidence interval Lower	Upper
Fixed effect	36	0.550	35	397.980	91.2	0.463	0.637
Random effects	36	0.842				0.540	1.144

Table 2 shows a fixed weighted average effect of g=0.550 and a random weighted average effect of g=0.842. Both the fixed and random weighted effect sizes are significantly greater than zero. The effect size is considered large by Cohen's standards. The Q statistics show that the distribution is significantly heterogeneous, and I-squared indicates that over 75% of variability in the distribution is between-study variance. Namely, variability in effect sizes exceeds sampling error. To explain this heterogeneity, moderator analysis was carried out.

In order to find an answer to the second research question, 'how do various alternative assessment techniques moderate the overall weighted effect size?', the included studies were classified into five categories in terms of alternative assessment techniques, such as peer assessment, self- and peer assessment, grid, portfolio and DBT and SG (diagnostic branched tree and structured grid). In accordance with these categories, the findings are presented in Table 3.

Table 3

Moderator Analysis of Various Alternative Assessment Techniques (AAT)

Variable	k	Effect size	95 % Confidence Interval Lower	Upper	Q_B	df	p
AAT	35				2.241	5	0.210
Peer assessment	2	0.423	-0.998	1.844			
Self- and peer assessment	3	0.877	-0.268	2.023			
Grid	2	0.629	-0.736	1.994			
Portfolio	24	1.012	0.604	1.420			
DBT and SG	4	0.501	-0.481	1.482			

As seen in Table 3, the learning environments in which portfolios are used have the largest effect size (d=1.012), and those with peer assessment have the smallest effect size (d=0.423). As Q value is smaller than the critical value ($Q_b < \chi^2$; $p>.05$), the Q-between is not significant for this variable, indicating that within chance they are equal.

The third research question, 'how do demographic study features moderate this effect size?' was formed to determine whether there is a significant difference between the effect sizes in terms of subject matter and study level. For subject matter analysis, some studies were excluded in this analysis, especially in subjects such as Computer, Chemistry, Social Science and Environmental Science, on which there are fewer studies. The findings are shown in Table 4.

Table 4

Moderator Analysis of Demographic Study Features

Variables	k	Effect size	95 % Confidence Interval		Q_B	df	p
			Lower	Upper			
Subject matter	29				2.661	2	0.264
Science and Technology	20	0.505	0.260	0.751			
English	6	0.861	0.409	1.313			
Math	3	0.905	0.251	1.559			
School level	36				26.069	2	0.000
Primary	23	0.549	0.176	0.922			
Secondary	4	3.137	2.205	4.069			
Undergraduate	9	0.648	0.059	1.237			

According to the findings in Table 4, the studies conducted in Math demonstrated the largest effect size (d=0.905), and those in Science and Technology showed the smallest effect (d=0.505). However, as the Q statistical value indicates, the distribution of effect sizes is found to be homogenous. In other words, there is no significant difference in effect size in terms of subject matter (Q_B=2.661; p= 0.264). The findings concerning school level show that the largest effect has been found at the secondary level (d=3.137), while the smallest effect is at the primary level (d=0.549). As the Q value exceeds the critical value with two degree of freedom, the distribution of the effect sizes is seen as heterogeneous (Q_B=26.069, p=0.000). Accordingly, the effect of alternative assessment techniques on academic achievement significantly varies by school level.

Publication Bias

In order to examine the publication bias, a funnel plot was drawn in Figure 3. As seen below, it is generally symmetrical around the mean of distribution. Accordingly, there is no publication bias comprising the results of this meta-analytic review. To

support this, upon analyzing Rosental's fail-safe N, it has been found out that the fail-safe N is 2048, based on 36 effect sizes from 26 studies, with a z value of 14.90 and corresponding p-value of 0.00. What this means is that this analysis must include 2048 'null' studies for p-value to exceed .05; that is, 56.8 missing studies would be required for each effect size to equate to 'zero'.

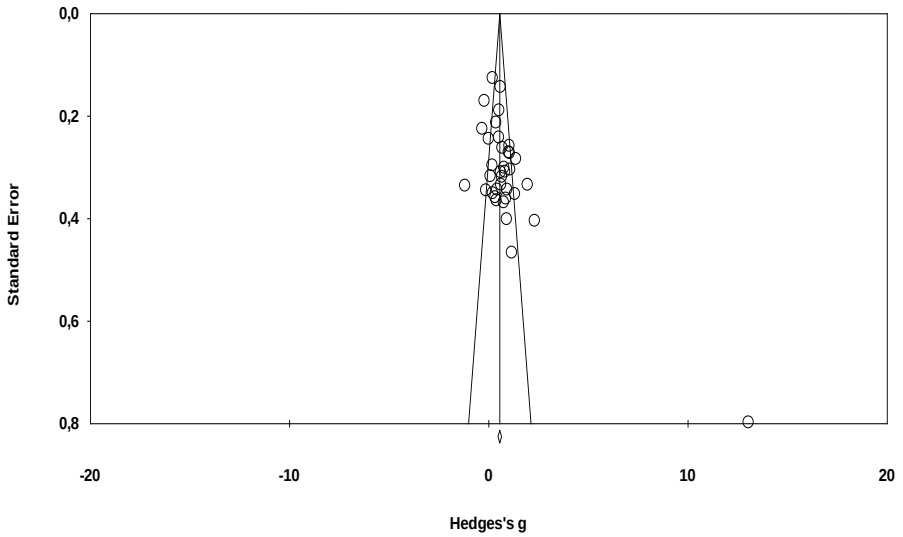

Figure 3. Funnel plot with effect sizes (horizontal axis) and standard errors (vertical axis)

Discussion and Conclusion

In accordance with 36 effect sizes derived from 26 studies conducted in Turkey, it has been revealed that AAT has a positive impact on academic achievement, and this effect has been classified as large by Cohen's standards. It has been concluded that AAT is significantly impact on student achievement. This result is suggestive enough, in addition to being congruent with many studies (Anahtarci, 2009; Bagci, 2009; Baris, 2011; Barootchi and Keshavarz, 2002; Fenwick and Parsons, 1999; Gungor, 2005; Gurel, 2013; Guven and Aydogdu, 2009; Izgi, 2007; Kirikkaya and Vurkaya, 2011; Koroglu, 2011; Memis, 2011; Menevse, 2012; Olgun, 2011; Ozek, 2009; Parlakyildiz, 2008; Turan, 2013).

This meta-analysis examined whether the estimated effect size varies in terms of various alternative assessment techniques, subject matter and school level. Moderator analysis of various alternative assessment techniques revealed that the studies conducted using portfolios in class have the largest effect, the ones using self- and peer assessment combined have a larger effect, and those using only peer assessment technique have the smallest effect. However, the results show that the effect of various alternative assessment techniques insignificantly varies. As a consequence of this

meta-analysis, it has been found that portfolios are most frequently used in the primary grades and represent a larger contribution to the weighted average effect size than the other techniques (Anahtarci, 2009; Gungor, 2005; Karamanoglu, 2006; Mihladiz, 2007; Okcu, 2007; Ozek, 2009; Parlakyildiz, 2008). In this sense, it may be considered that the effect of other techniques on achievement is of importance and should be further explored.

The other moderator analysis was carried out on descriptive subject features such as subject matter and school level in which the primary studies were conducted. In terms of subject matter, the results demonstrated that the treatments in Mathematics courses have larger effect size, while those in Science and Technology courses have relatively low effect size. However, based on the findings, it has been stated that the effect of AAT on achievement does not differ in terms of subject matter. As for school level, the results show that interventions in the secondary schools have a large effect size, whereas those in the primary schools have a moderate effect size. On the other hand, it has been revealed that there is a significant difference in effect sizes in terms of school level, and the impact of AAT on achievement differs with regard to the school level. Winking (1997) stated that alternative assessment requires upper cognitive skills, so students can solve real-life problems. Additionally, it is known that what is effective in alternative assessment is that critical thinking and creativity develop over time (Eva, Cunnington, Reiter, Keane and Norman, 2004).

It is essential that meta-analytical results be interpreted with consideration to some of the limitations of primary studies. Some factors such as the experiment period, the experimenter's characteristics, and the difficulties in the experiments likely affect the results. According to Corcoran, Dershimer and Tichenor (2004), even though many teachers agree on the importance of using any kind of alternative assessment techniques, they state that it is difficult to administer them to the students.

In the current meta-analysis, the effect of alternative assessment techniques has been examined only in regard to student academic achievement. The effect of AAT on attitudes, anxiety and motivation may be investigated in future meta-analytic studies. In the literature review for this meta-analysis, it has been noted that there is a lack of study in some subject matter areas. Accordingly, the comparison of effect sizes in terms of subject matter has fallen short. More experimental / quasi-experimental studies may be conducted in other subject matters such as Turkish Language, History, and Chemistry. Considering the limited studies on AAT conducted in Turkey, a new meta-analytic study may be designed, including the studies on AAT from other countries.

References

(The studies with asterisk indicate that the studies are included in this review.)

*Anahtarci, S. (2009). *Yapilandirmaci yaklasima dayali matematik programinda portfolyonun basariya ve matematige karsi tutuma etkisi*. [The influence of the portfolio on success and mathematics in a constructivist approach-based mathematics program] Yayimlanmamis yuksek lisans tezi. Trakya Universitesi Sosyal Bilimler Enstitusu, Edirne. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi.

*Bagci, I. (2009). *Ilkogretim sosyal bilgiler dersinde Turk tarihinde yolculuk unitesinin alternatif degerlendirme (portfolyo) ile islenmesinin ogrenci basarisi ve tutumuna etkisi*.[The effects of student achievement and attitude of the unit 'voyage in the history of Turkish' with the evaluation of alternative processing (portfolio) in the lesson of social studies courses] Yayımlanmamış yüksek lisans tezi. Gazi Universitesi Egitim Bilimleri Enstitusu, Ankara. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Balaban, M. (2010). *Ilkogretimde fen ve teknoloji dersinde ogretmen, ogrenci, veli isbirligi ile portfolyo uygulamasi.* [Teachers, students, parents cooperation in ensuring portfolio assessment in elementary school for science and technology lesson]. Yayimlanmamis yuksek lisans tezi. Ondokuz Mayıs Universitesi Fen Bilimleri Enstitusu, Samsun. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Baris, M. F. (2011). *Bir sosyal ag sitesine e-portfolyonun (elektronik gelisim dosyasi) entegre edilerek uygulanmasi ve sonuclarinin incelenmesi*. [Integration of the e-portfolio into a social network and analysis of results]. Yayimlanmamis doktora tezi. Trakya Universitesi Fen Bilimleri Enstitusu, Edirne. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Barootchi, N., & Keshavarz, M. H. (2002). Assessment of achievement through portfolios and teacher-made tests. *Educational Research*, 44 (39), 279–288.

Buyuktokatli, N., & Bayraktar, S. (2014). Fen egitiminde alternatif olcme degerlendirme uygulamalarıı. [Alternative assessment applications in science education] *Pegem Egitim ve Ogretim Dergisi, 4*(1), 103-126.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98-101.

Cooper, H. (2010). *Research synthesis and meta-analysis (A Step by Step Approach)* 4th edition. Thousand Oaks: SAGE Publication

Corcoran, A. C., Dershimer, L. E. & Tichenor, S. M. (2004). A teacher's guide to alternative assessment –taking the first steps. *The Clearing House*, 77 (5), 213–216.

Coruhlu, T.S., Er Nas, S. & Cepni, S. (2009). Fen ve teknoloji ogretmenlerinin alternatif olcme degerlendirme tekniklerini kullanmada karsilastiklari problemler:

Trabzon ornegi [Problems faced by science and technology teachers in using alternative assessment and evaluation techniques: Trabzon example].*Yuzuncu Yil Universitesi Egitim Fakultesi Dergisi,* V1 (1), 122-141.

*Dogan, S. (2012). K*avram haritasi ve yapilandirilmis grid tekniginin coktan secmeli testlerle karsilastirilmasi.* [A comparison of concept map and structural grid techniques with multiple choice tests]. Yayimlanmamis yuksek lisans tezi. Mersin Universitesi Eğitim Bilimleri Enstitusu, Mersin. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Dos, B. (2016). Analyzing the alternative assessment applications for the development of teaching: review of literature. *International Online Journal of Educational Sciences*, 8 (4), 215-228.

Duban, N. & Kucukyilmaz, E.A. (2008). Sınıf ogretmeni adaylarinin alternatif olcme degerlendirme yontem ve tekniklerinin uygulama okullarında kullanımına iliskin gorusleri. [Primary school student teachers' opinions regarding the use of alternative assessment methods and techniques in practice schools]. *Ilkogretim Online, 7*(3), 769-784.

*Erdogan, T. (2006). *Yabanci dil ogretiminde portfolyoya dayali degerlendirmenin ogrenci basarisi ve derse yonelik tutumlarina etkisi.* [Effect of portfolio-based assessment on students' achievement and attitudes in foreign language teaching]. Yayımlanmamis yuksek lisans tezi. Dokuz Eylul Universitesi Egitim Bilimleri Enstitusu, İzmir. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Erkus, A. (2006). *Sınıf ogretmenleri için olcme ve degerlendirme kavramlar ve uygulamalar.* [Concepts and practices on measurement and evaluation for primary school teachers] Ankara: Ekinoks Yayınları.

Eva, K., J.P.W. Cunnington, H. I. Reiter, D. R. Keane, & Norman, G. R. (2004). How can I know what I don't know? Poor self-assessment in a well-defined domain. *Advances in Health Sciences Education*, 9, 211–224

Fenwick, T. J. & Parson, J. (1999). A note on using portfolios to assess learning. *Canadian Social Studies*, 33(3), 90-92.

Gelbal, S. & Kelecioglu, H. (2007). Ogretmenlerin olcme ve degerlendirme yontemleri hakkındaki yeterlik algıları ve karsılastıkları sorunlar**.** [Teachers' perception of competence about the measurement and evaluation methods and problems they encounter] *Hacettepe Universitesi Egitim Fakultesi Dergisi,* 33, 135-145.

Gerek, O. (2006). *Sınıf ogretmenlerinin yeni ilkogretim programı hakkındaki gorus degerlendirme ve yeterlilikleri uzerine bir arastirma.* [A study on the primary school teachers' opinions, competences and evaluations about the new primary education program] Yayımlanmamis yuksek lisans tezi. Selcuk Universitesi Sosyal Bilimler Enstitusu, Konya. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Greenstein, L. (2010). *What teachers really need to know about formative assessment*. Alexandria, Virginia USA: ASCD.

Gummer, S. & Shepardson, D.P. (2001). Facilitating change in classroom assessment practice: issues for professional development. In Shepardson D.P. (ed.), *Assessment in science* (pp. 53-66) Springer, Dordrecht.

*Gungor, S. (2005). *Ortaogretim geometri dersi ucgenler konusunda oluşturmacı (constructivism) yaklaşıma dayalı elle yapılan materyaller ve portfolyo (portfolio) hazırlamanın ogrenciler uzerindeki etkilerinin incelenmesi.* [Examining the effects of preparing portfolio and hands on materials which is based on the constructivist approach, on students in instructing triangles subject in geometry lesson] Yayımlanmamış yüksek lisans tezi. Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Enstitusu, Zonguldak. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Gurel, H. (2013). *İlköğretim 7. ve 8.sınıf fen ve teknoloji dersinde portfolyo uygulamasının öğrencilerin akademik başarı ve hatırlama düzeyine etkisi.* [The effect of using portfolio over students' academic success and remembering level for 7th and 8th grades for Science and Technology lesson at primary schools]Yayımlanmamış yüksek lisans tezi. Çanakkale Onsekiz Mart Üniversitesi Eğitim Bilimleri Enstitüsü, Çanakkale. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Guven, E. & Aydoğdu, M. (2009). Portfolyonun 6. sınıf fen ve teknoloji dersi vucudumuzda sistemler ünitesinde basarı ve kalıcılıga etkisi. [The effect of portfolio on the achievement and permanence in the unit 'system in our body' of 6th grade science and technology course] *Turk Fen Egitimi Dergisi, 6(2), 115-128.*

Hassan, A.K., Hill, R.A., & Reid, N. (2004). Ideas underpinning success in an introductory course in organic chemistry. *The Higher Education Chemistry Journal of the Royal Society of Chemistry*, 8 (2), 40-52.

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

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

*Izgi, U. (2007). *Fen egitiminde alternatif degerlendirme yaklasimlarının ogrencilerin sınav kaygısına ve ogrenmede kalıcılıga etkisi*.[The effect of alternative assessment attitudes in science lectures on the level of examination anxiety of students and the level of performance of learning] Yayimlanmamis yuksek lisans tezi. Hacettepe Universitesi Sosyal Bilimler Enstitusu, Ankara. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Johnstone, A.H., Bahar, M. & Hansell, M.H. (2000). Structural communication grids: a valuable assessment and diagnostic tool for science teachers. *Journal of Biological Education*, 34 (2), 87-89.

Kablan, Z.,Topan, B. & Erkan, B. (2013). Sınıf ici ogretimde materyal kullaniminin etkililik duzeyi: Bir meta-analiz calisması. [The effectiveness level of material use in classroom instruction: a meta-analysis study] *Kuram ve Uygulamada Egitim Bilimleri,* 13 (3), 1629-1644.

Karahan, U. (2007). *Alternatif olcme ve degerlendirme metotlarından grid, tanilayici dallanmis agac ve kavram haritalarının biyoloji ogretiminde uygulanması*, [Application of alternative measurement and evaluation methods that are grid, diagnostic tree and concept maps within biology education] Yuksek Lisans Tezi, Gazi Universitesi Egitim Bilimleri Enstitusu, Ankara. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Karakus, U.,& Ozturk Demirbas, Ç. (2011). Coğrafya öğretmenlerinin ölçme ve değerlendirme araçlarını kullanımı. [The use of measurement and evaluation tools by geography teachers] *Milli Eğitim Dergisi 189,* 71- 81.

Karakus, F. (2010). Ortaöğretim matematik dersi öğretim programında yer alan alternatif ölçme ve değerlendirme yaklaşımlarına yönelik öğretmen görüşleri. [Teacher views on alternative assessment and evaluation approaches in secondary school mathematics curriculum] *Türk Eğitim Bilimleri Dergisi, 8(2),* 457-488.

*Karamanoglu, S. S. (2006). *İlköğretim öğrencilerinin fen başarılarının değerlendirilmesinde sorgulama programının kullanılması: portfolyo.* [Assessing primary students success in science courses via primary years program: Portfolio] Yayımlanmamış yüksek lisans tezi. Marmara Üniversitesi Eğitim Bilimleri Enstitüsü, İstanbul. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Kaya, S. (2013). *Isbirlikli ogrenme ve akran degerlendirmenin akademik basari, bilisustu yeti ve yardim davranislarina etkisi*.[Effects of cooperative learning and peer assessment on academic success, metacognitive awareness and helping behaviors] Yayımlanmamış yuksek lisans tezi. Cumhuriyet Universitesi Egitim Bilimleri Enstitusu, Sivas. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Kirikkaya, E. B. & Vurkaya, G. (2011). The effect of using alternative assessment activities on students' success and attitudes in science and technology course. *Educational Sciences: Theory and Practice*, *11*(2), 997-1004.

*Koc, A. (2010). B*ilgisayar okuryazarligi becerisi ediniminde e-portfolyo surecinin ogrenen performansina ve tutumlarina etkisi*.[The effect of e-portfolio process on learners' performance and attitudes in the process of computer literacy skill acquisition] Yayımlanmamış yuksek lisans tezi. Fırat Universitesi Sosyal Bilimler Enstitusu, Elazıg. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Korkmaz, I. (2006). Yeni ilkogretim programinin ogretmenler tarafindan degerlendirilmesi. [Evaluation of new primary education program by

teachers] *Ulusal Sınıf Ogretmenligi Kongresi,* Cilt II (s.249-260). Ankara: Kök Yayıncılık.

Kosteklioglu, I. & Celen, U. (2016). Oz degerlendirme yonteminin etkililiginin degerlendirilmesi. [Evaluation of the effectiveness of self-assessment method]. *Elementary Education Online*, 15(2), 671-681.

*Koroglu, U. M. (2011). *İngilizce ogretiminde proje tabanlı ogrenme ve portfolyo degerlendirme uygulamalarının lise ogrencilerinin başarısına etkisi.* [The effects of project-based and portfolio based learning on high school students' reading and writing skills in English]. Yayimlanmamis yuksek lisans tezi. Uludag Universitesi Egitim Bilimleri Enstitusu, Bursa. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

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

*Memis, E. K. (2011). A*rgumantasyon tabanli bilim ogrenme yaklasiminin ve oz degerlendirmenin ilkogretim ogrencilerinin fen ve teknoloji dersi basarisina ve basarinin kaliciligina etkisi.* [Effects of the argumentation based science learning approach and self-evaluation on primary school students' science and technology course achievement and retention of the achievement] Yayimlanmamis doktora tezi. Ataturk Universitesi Egitim Bilimleri Enstitusu, Erzurum. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Menevse, E.B. (2012). *Portfolyo uygulamasinin ogretmen adaylarinin akademik basarilarina etkisi*. [Effects of teachers academic achievement application portfolio]. Yayimlanmamis yuksek lisans tezi. Necmettin Erbakan Universitesi Egitim Bilimleri Enstitusu, Konya. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Mihladiz, G. (2007). *Ilkogretim fen bilgisi ogretiminde portfolyo uygulamasinin ogrencilerin akademik basarilarina ve derse yonelik tutumlarina etkisi.* [The effect of application of portfolio in primary school science teaching on the students' academic achievements and attitudes towards the lesson] Yayimlanmamis yuksek lisans tezi. Mugla Universitesi Fen Bilimleri Enstitusu, Mugla. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

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

MEB (Milli Egitim Bakanligi) (2005). *Ilkogretim Turkce dersi ogretim programi*. [The curriculum of primary school Turkish course] Ankara: MEB Yayinlari.

*Okcu, Y. (2007). *Matematik egitiminde portfolyo degerlendirme.* [Portfolio assessment in mathematics education]. Yayimlanmamis yuksek lisans tezi. Balikesir Universitesi Fen Bilimleri Enstitusu, Balıkesir.

*Olgun, M. (2011). *Ilkogretim 4. sinif fen ve teknoloji dersinde oz ve akran degerlendirme uygulamalarinin yer aldigi isbirlikli ogrenme yonteminin ogrencilerin basari, tutum*

ve bilisustu becerilerine etkisi. [The effect of cooperative learning method including self and peer assessment applications on student's success, attitude and metacognitive skills in primary 4th grade science and technology lesson]. Yayimlanmamis yuksek lisans tezi. Dokuz Eylul Universitesi Egitim Bilimleri Enstitusu, Izmir. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Ozcan, F. (2011). *9. sinif ogrencilerinin kimyasal degisimler konusundaki kavramsal basarilari uzerine alternatif degerlendirme tekniklerinin etkisi.* [The effect of the alternative assessment techniques to first grade of high school students' conceptual success on the subject of the chemical changes]. Yayimlanmamis yuksek lisans tezi. Gazi Universitesi Egitim Bilimleri Enstitusu, Ankara. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Ozdemir, S.M. (2010). Ilkogretim ogretmenlerinin alternatif olcme ve degerlendirme araclarina iliskin yeterlikleri ve hizmet içi egitim ihtiyaclari. [Competencies of elementary school teachers on alternative assessment tools and in-service training needs].*Turk Egitim Bilimleri Dergisi,* 8(4), 787-816.

*Ozek, S. (2009). *Portfolyo uygulamasinin ilkogretim ogrencilerinin Ingilizce dersinde okudugunu anlama duzeylerine ve derse yonelik tutumlarina etkisi.* [The effect of portfolio application on primary school students? reading comprehension skills in English and their attitudes towards English]. Yayimlanmamis yuksek lisans tezi. Ankara Universitesi Egitim Bilimleri Enstitusu, Ankara. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Ozsevgec, T., & Karamustafaoglu, S. (2010). Ogretmen adaylarinin geleneksel ve yapilandirmaci olcme-degerlendirme yaklasimlarina yonelik profilleri. [Profiles of teacher candidates for traditional and constructive assessment and evaluation approaches]. *Turk Egitim Bilimleri Dergisi,* 8(2), 333-354.

*Parlakyildiz, B. (2008). *Portfolyoya dayali degerlendirmenin universite ogrencilerinin akademik basarilarina ve bilissel yasam becerilerine etkisi.* [The impact of portfolio-based assessment on the academic accomplishment and cognitive life skills of university students] Yayımlanmamış doktora tezi. Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

*Turan, M. A. (2013). *Ilkogretim besinci sinif fen ve teknoloji dersinde portfolyo uygulamasinin ogrenci basarisi ve ogrenmenin kaliciligi ızerine etkisi.*[The effectiveness of the portfolio method on the students' academic success and retention outcomes in fifth-grade science classrooms] Yayimlanmamis yuksek lisans tezi. Marmara Universitesi Egitim Bilimleri Enstitusu, Istanbul. Retrieved from https://tez.yok.gov.tr/UlusalTezMerkezi

Winking, D. (1997). *Critical issue: Ensuring equity with alternative assessments* [online document]. NCREL (North Central Regional Educational Laboratory), Oak Brook: IL. Retrieved January 15, 2015, from http://www.ncrel.org/sdrs/areas/methods/assment/as800.htm

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. London: Sage.

Yaman, S. (2011). Teachers' perceptions about their measurement and evaluation practices in science and technology course, *İlköğretim Online*, 10(1), 244-256.

Yazici, F. & Sozbilir. M. (2014). Elementary 6-8 grades teachers' frequency of use and their level of adequacy in assessment and evaluation methods: Erzurum sampling. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, (8), 2, 164-196.

Yesilyurt, E. (2012). Fen ve teknoloji dersinde kullanilan olcme degerlendirme yontemleri ve karsilasilan guclukler. [Measurement and evaluation methods used in science and technology lessons and difficulties encountered] *Turkish Studies - International Periodical for the Languages, Literature and History of Turkish or Turkic*, 7(2), 1183-1205.

Alternatif Değerlendirme Tekniklerinin Etkililiğinin Meta-Analizi

Atıf:

Gozuyesil, E. & Tanriseven, I. (2017). A Meta-analysis of the effectiveness of alternative assessment techniques. *Eurasian Journal of Educational Research*, 70, 37-56, DOI: http://dx.doi.org/10.14689/ejer.2017.70.3

Özet

Problem Durumu: Öğretmenlerin, öğrencilerini ve kendi öğretim faaliyetlerini değerlendirmede öğretim programlarında tavsiye edilen ölçme değerlendirme tekniklerini kullanmaları önem taşımaktadır. Türkiye'de 2004 eğitim reformu ile yenilenmekte olan öğretim programları öğrencilerin, çok yönlü ve üst düzey düşünme becerilerini açığa çıkarıcı bir şekilde değerlendirilmesi gerektiğini öne sürmekte ve geleneksel değerlendirme yöntemleriyle birlikte alternatif ölçme ve değerlendirme tekniklerinin kullanılmasını önermektedir. Bu ölçme araçlarının etkililiği programa yabancı olan ve bu araçları kullanan öğretmenler tarafından şüphe ile karşılanmakta ve bu durumda araçların uygulanması da oldukça güçleşmektedir. Alternatif değerlendirme tekniklerinin etkililiği ile ilgili günümüze kadar pek çok araştırma yapıldığı ve hala yapılmakta olduğu görülmektedir. Fakat hem alternatif değerlendirme tekniklerinin etkililiğini geniş ölçekte ortaya koyan hem de hangi değerlendirme tekniğinin akademik başarı üzerinde daha etkili olduğunu gösteren bir çalışmaya alan yazında rastlanmamıştır. Bu çalışma, son on yılda ülkemizde yapılan program değişikliği sebebiyle popülerliği gittikçe artan alternatif değerlendirme ile ilgili alan yazını gözden geçirmek için planlanmıştır. Bireysel araştırmalardan elde edilen veriler ve bulguların meta-analiz yöntemiyle birleştirilmesi araştırmanın temelini oluşturmaktadır. Bu sayede alternatif yöntemlerle değerlendirilen akademik

başarının etki büyüklüğüne ulaşılması ve hangi değerlendirme tekniğinin daha etkili olduğunu tartışmaya imkân sağlayabilecektir.

Araştırmanın Amacı: Bu araştırma, alternatif değerlendirme tekniklerinin öğrencilerin akademik başarılarına etkisini ve akademik başarının kullanılan alternatif değerlendirme tekniği türlerine, tekniğin uygulandığı ders türüne ve öğretim kademesine göre farklılaşıp farklılaşmadığını meta-analiz yöntemiyle araştırmayı hedeflemiştir.

Araştırmanın Yöntemi: Bu çalışmada alternatif değerlendirme tekniklerinin akademik başarı üzerindeki etkililiği üzerine yapılmış birincil çalışmaların etki büyüklüklerinin hesaplanması, birleştirilmesi ve yorumlanması amacıyla meta-analiz yöntemi kullanılmıştır. Bu anlamda öncelikle ilgili birincil çalışmalara ulaşmak için Türkiye'de alternatif değerlendirme tekniklerinin eğitim programlarına dâhil edildiği yıl olan 2004 yılı itibariyle yapılmış çalışmaların alan yazın taraması yapılmıştır. Bu taramada "alternatif değerlendirme", "portfolyo", "yapılandırılmış grid", "tanılayıcı dallanmış ağaç", "akran değerlendirme", "öz değerlendirme" gibi anahtar sözcükler ile YÖK Ulusal Tez Merkezi, ERIC, PsycINFO, ASOS sosyal bilimler indeksi gibi veri tabanları ile üniversitelerin eğitim fakültesi dergileri taranmıştır. İlgili alan yazın çalışmasından sonra (1) alternatif değerlendirme tekniğinin öğrencilerin akademik başarısı üzerindeki etkisini inceleyen, (2) ön-test son-test deneysel ya da yarı deneysel en az iki bağımsız örneklem içeren, (3) etki büyüklüğü hesaplamaları için gereken istatistiksel veriler içeren, (4) alternatif değerlendirme tekniğini Türkiye'de uygulamış olan ve (5) 2004 – Ağustos 2014 yılları arasında yapılmış olan 26 çalışma analize dâhil edilmiştir. Bu çalışmalar araştırmacı tarafından geliştirilen ve % 98 oranında değerlendiriciler arası güvenirliğin tespit edildiği bir kodlama formuna işlenmiştir. . Bu çalışmada her çalışmanın etki büyüklüğü hesaplandıktan sonra etki büyüklüklerinin homojenliği testi için Cochran tarafından önerilen (k-1) serbestlik dereceli Ki-Kare heterojenlik testi (Q istatistiği) kullanılmıştır. Heterojenlik derecesinin belirlenmesi için ise I^2 testi seçilmiştir. Meta-analizde sabit etkiler modeli uygulandığında elde edilen; grup içi, gruplar arası ve toplam heterojenlik değerlerinin kritik değerlerden yüksek olduğu ortaya çıkmıştır. Bu nedenle rastgele etkiler modeli kullanılarak etki büyüklükleri tekrar hesaplanmıştır. Heterojenliğin kaynağının tespiti için ise bazı kategorik değişkenler için moderatör analizi yapılmıştır. Yayın yanlılığın olup olmadığının test edilmesi için huni diyagramı oluşturulmuş ve Rosental'ın korumalı N testi ile sonuç desteklenmiştir.

Araştırmanın Bulguları: Çalışmaların etki büyüklüklerinin heterojen yapıda ($Q>x^2$, $p< 0.05$) ve çalışmalar arasındaki heterojenlik miktarının ($I^2=91$) yüksek çıkması sonucunda yapılan moderatör analizi heterojenlik kaynağının birincil çalışmaların yapıldığı öğrenim kademeleriyle ilişkili olduğu görülmüştür. Diğer bir deyişle etki büyüklükleri çalışmaların yapıldığı ders türlerine göre ve çalışmalarda kullanılan alternatif değerlendirme teknikleri türlerine göre farklılaşmazken öğrencilerin öğrenim kademelerine göre etki büyüklükleri arasında anlamlı bir farklılık olduğu ortaya çıkmıştır. Bulgular alternatif değerlendirme tekniklerinin öğrencilerin akademik başarısı üzerinde pozitif ve yüksek düzeyde bir etkiye (d=0.84) sahip

olduğunu ortaya koymuştur. Ayrıca, bu tekniklerin kullanılması öğrencilerin matematik dersindeki akademik başarıları üzerinde geniş bir etkiye sahip olduğu (.90) ve portfolyo kullanımının da (d=1.01) etkisinin kayda değer olduğu sonucuna ulaşılmıştır. Yapılan yayın yanlılığı analizi sonucunda, bu meta-analiz bulgularını çarpıtacak bir yanlılığın olmadığı elde edilen değerlerin yüksek güvenirlikte olduğu belirlenmiştir.

Araştırmanın Sonuçları ve Önerileri: Mevcut araştırmanın sonuçları, alternatif değerlendirme tekniklerinin öğrencilerin akademik başarısı açısından geleneksel değerlendirme tekniklerine oranla daha başarılı olduğunu göstermiştir. Buna ek olarak, alternatif değerlendirmenin etkililiği kullanılan değerlendirme tekniklerine ve ders türlerine göre anlamlı farklılık göstermediğinden farklı otantik değerlendirme tekniklerinin derslerde kullanılması öğrenci başarısını artırdığı söylenebilir. Ancak, bu tekniklerin etkililiği kullanıldığı öğrenim kademelerine göre anlamlı fark gösterdiği sonucuna ulaşılmıştır. Alternatif değerlendirmeler daha üst düşünme becerileri gerektirdiği için alt kademlerde kullanımında daha az etkili olduğu sonucuna varılmıştır. Bu tekniklerin özellikle ortaöğretim ya da yükseköğretimde kullanılmasının daha etkili olacağı söylenebilir. Bu meta-analiz çalışmada alternatif değerlendirme tekniklerinin etkililik düzeyi sadece akademik başarı açısından incelenmiştir. Yapılacak diğer meta-analiz çalışmalarında alternatif değerlendirme tekniklerinin tutum, kaygı ve motivasyon düzeyi açısından etkililiği araştırılabilir. Çalışmada meta-analiz için yapılan alan taramasında bazı derslere yönelik yeterli çalışmaya ulaşılamamıştır. Dolayısıyla ders türlerine göre karşılaştırmalarda bazı boyutlar eksik kalmıştır. Buna göre diğer dersler için yeni araştırmalar yapılması alan yazına katkı sağlayabilir.

Anahtar Kelimeler: Otantik değerlendirme, portfolyo, performans, etki büyüklüğü.