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

Eda GOZUYESI1, Isil TANRISEVEN2

**ARTICLE INFO**

**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.

© 2017 Ani Publishing Ltd. All rights reserved

---

*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, isiltanrisevent77@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 multidimensionally 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; Kosterelioglu 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.
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 \( d = \frac{X_e - X_c}{SD} \) (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

**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 \( d = \frac{X_e - X_c}{SD} \) (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. F 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

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

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

<table>
<thead>
<tr>
<th>Analytical models</th>
<th>N</th>
<th>Effect size</th>
<th>df</th>
<th>Q_total</th>
<th>I²</th>
<th>95% Confidence interval</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect</td>
<td>36</td>
<td>0.550</td>
<td>36</td>
<td>397.980</td>
<td>91.2</td>
<td>0.463</td>
<td>0.637</td>
<td></td>
</tr>
<tr>
<td>Random effects</td>
<td>36</td>
<td>0.842</td>
<td>35</td>
<td>91.2</td>
<td>91.2</td>
<td>0.540</td>
<td>1.144</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>k</th>
<th>Effect size</th>
<th>Q Total</th>
<th>I²</th>
<th>95% Confidence interval</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer assessment</td>
<td>2</td>
<td>0.423</td>
<td>-0.998</td>
<td>1.844</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self- and peer assessment</td>
<td>3</td>
<td>0.877</td>
<td>-0.268</td>
<td>2.023</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grid</td>
<td>2</td>
<td>0.629</td>
<td>-0.736</td>
<td>1.994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portfolio</td>
<td>24</td>
<td>1.012</td>
<td>0.604</td>
<td>1.420</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBT and SG</td>
<td>4</td>
<td>0.501</td>
<td>-0.481</td>
<td>1.482</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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>0.05), 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

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

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; Kırıkkaya and Vurkaya, 2011; Koroglu, 2011; Memis, 2011; Menevse, 2012; Olgun, 2011; Ozek, 2009; Parlakıylı, 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; Parlayildiz, 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.)


Alternatif Değerlendirme Tekniklerinin Etkililiğinin Meta-Analizi

Atıf:

Özet
Araştırmanın Amacı: Bu araştırma, alternatif değerlendirme tekniklerinin öğrencilerin akademik başarılara etkisini ve akademik başarının kullanılan alternatif değerlendirme teknikleri türlerine, teğkinin uygulanındaki ders türine ve öğretim kademesine göre farklılaşıp farklılaşmadığını meta-analiz yöntemiyle araştırmayı hedeflemiştir.


Araştırmanın Bulguları: Çalışmaların etki büyüklüğünün heterojen yapıda (Q>χ², p< 0.05) ve çalışmalar arasındaki heterojenlik miktarının (I=91) yüksek olması sonucunda yapılan moderatör analizi heterojenlik kaynağındaki birincil çalışmaların yapıldığı ögrenim kademesiyle ilişkili olduğu görülmüşür. Diğer bir deyişle etki büyüklükleri çalışmalara yapıldığı ders türlerine göre ve çalışmalarda kullanılan alternatif değerlendirme teknikleri türlerine göre farklılaşmışken öğrencilerin öğrenim kademesine göre etki büyüklükleri arasında anlamlı bir farklılık olduğu ortaya çıkmıştır. Bulgular alternatif değerlendirme tekniklerinin öğrencilerin akademik başarı üzerinde pozitif ve yüksek düzeyde bir etkiye (d=0.84) sahip
olduğunu ortaya koymuştur. Ayrıca, bu tekniklerin kullanması öğrencilerin matematik dersindeki akademik başarısını üzerinde geniş bir etkiye sahip olduğunu (.90) ve portfolyo kullanmanın da (d=1.01) etkisinin kayda değer olduğunu sonucuna ulaşmıştır. Yapılan yayın yanlılığı analizi sonucunda, bu meta-analiz bulgularını çarpıtacak bir yanlılığın olmadığı elde edilen değerlerin yüksek güvenilirlikte olduğu belirlenmiştir.


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