

The Effect of Formative Assessment Practices on Student Learning: A Meta-Analysis Study

Pinar Karaman ^{1,*}

¹Sinop University, Faculty of Education, Department of Educational Sciences, Sinop, Turkey

ARTICLE HISTORY

Received: Jan. 28, 2021

Revised: Aug. 22, 2021

Accepted: Sep. 22, 2021

Keywords:

Formative assessment,
Student learning,
Meta-analysis.

Abstract: The main purpose of this meta-analysis study is to investigate how formative assessment practices promote student learning in Turkey. 32 studies with 47 effect sizes that met the specified criteria such as using true experimental or quasi-experimental design and measuring learning outcomes were included as the final analysis in the meta-analytical review method. The overall mean effect size of the study was obtained as .72 ($SE = .07, p < .05$). Further investigation through subgroup analysis showed that the effect sizes made a significant difference on different types of formative feedback. The effect of features of formative assessment interventions on student learning indicated that student initiated formative feedback ($d = 1.16$) and mixed feedback ($d = .83$) had a large effect, which was followed by a medium effect of adult initiated formative feedback ($d = .69$) and a small effect of computer initiated formative feedback ($d = .42$). On the other hand, education level and publication type had no effect on student academic performance in the study. These findings support the positive effect of formative assessment practices on student learning. Such a result suggests that increasing the number of different types of formative assessment practices in the classrooms would promise a considerable contribution to student learning.

1. INTRODUCTION

Assessment is an important component of effective teaching and learning (Bransford et al., 2000; Hargreaves, 2008). Formative assessment strategy plays a crucial role in supporting the student learning. This assessment strategy provides effective feedback and instructional correctives in the teaching-learning process to improve students' learning, motivation, and self-regulation skills (Black & William, 2009; Cauley & McMillan, 2010; McManus, 2008; Popham, 2008). Formative assessment also known as assessment for learning, diagnostic testing, and feedback is an ongoing process used by teachers, students, and students' peers (Andersson & Palm, 2017a, 2017b; Bennett, 2011). Teachers can adjust their teaching practices to increase student learning through formative assessment (Black & William, 2009; Brookhart, 2009).

Formative assessment has the succeeding three main stages; namely, (1) determining goals, (2) providing feedback to enhance student performance with these goals, and (3) using feedback to

*CONTACT: Pinar KARAMAN ✉ pkaraman@sinop.edu.tr 📍 Sinop University, Faculty of Education, Department of Educational Sciences, Sinop, Turkey

improve further learning of students (Brookhart, 2010). One of the most important components of formative assessment is feedback that helps to provide evidence on student learning (Andersson & Palm, 2017a). Feedback helps students to understand current status of their learning to make further progress (Sadler, 1989). This feedback to advance student learning could come from different agents such as teachers, self-assessment, peer assessment, group assessment, and even computers (Sadler, 1989; Black & William, 1998; Graham et al., 2015; Wiliam, 2018). Feedback may be given to students in different time periods (instantly or delayed) (Andersson & Palm, 2017a). Thus, different types of feedback provide different formative assessment interventions (Hattie & Timperley, 2007). Feedback from teachers and students has an important role in formative assessment practices due to their significant support for student learning (Black & William, 2009), self-regulated learning (Andrade & Brookhart, 2016; Butler & Winne, 1995; Zimmerman & Bandura, 1994), and peer-assisted learning (Gielen et al., 2010). Students' engagement in self-assessment and peer-assessment for effective formative assessment strategies promotes their self-regulated learning skills (Zimmerman, 2002; Weldmeskel & Michael, 2016). In addition to teacher and student initiated formative assessment, computer initiated formative assessment also provides immediate feedback to students (Maier et al., 2016; Van der Kleij et al., 2015). These studies showed that computer-based feedback has an important effect on student learning (Kluger & DeNisi, 1996; Miller, 2009). However, in comparison to formative feedback from teachers and students, computer-based formative assessment is more difficult to apply (Maier et al., 2016).

Several meta-analysis research studies have been conducted to investigate the efficiency of formative assessment strategies. The results of these studies indicate that effect sizes vary with a considerable range (Black & William, 1998; Fuchs & Fuchs, 1986; Graham et al., 2015; King & Nash, 2011; Lee et al., 2020). The magnitude of the effect sizes of differences could come from a variety of the meta-analysis studies that focused on formative assessment types, feedback procedures, and learning subjects. (Maier et al., 2016). Black and William (1998) provided meta-analysis of 250 studies on the effect of formative assessment practices and found positive influence of formative assessment on student achievement with effect sizes ranging from .40 to .70. They argued that formative assessment intervention is more important than other educational interventions to improve student learning. Hattie (2009) examined the factors that were significantly related to student achievement through multiple meta-analysis and found that one of the most important factors is teachers' use of formative assessment strategies. Kingston and Nash (2011), in their meta-analysis research, examined a limited number of studies (a total of 13 studies) to uncover the effect of formative assessment on K-12 student achievement and reported the mean effect size as .20. They suggested that more studies are needed to investigate the relationship between formative assessment and academic achievement. On the other hand, Graham et al. (2015) investigated the effect of formative assessment on students' writing performance and reported a weighted mean effect size of .61. They also reported the impact of feedback from adults ($d=.87$), feedback from students (peer assessment, $d=.58$ and self-assessment, $d=.62$), and feedback from computers ($d=.38$) to student writing performance. Lee et al. (2020) analyzed 33 studies about K-12 education in the USA and reported an overall mean effect size of .29. They found the effectiveness of formative assessment on different subject areas. Moreover, meta-regression analysis denoted that student-initiated self-assessment was the most effective one ($d=.61$) among other interventions. In comparison to informal feedback ($d=.52$), formal formative assessment feedback was more effective on student learning. Briefly, although several meta-analysis studies in the literature concluded that formative assessment has a positive effect on student learning, the effectiveness of different types of formative assessment interventions was not examined adequately in previous meta-analysis studies.

For more than a decade, Turkey has given priority to the improvement of assessment for learning in education programs and offered more support to teachers to encourage them to use this assessment strategy more frequently in their classrooms (Kitchen et al., 2019; MoNE, 2017, 2020). With the growing importance of using formative assessment strategies in classrooms, the number of research studies conducted about the effectiveness of the formative assessment has increased considerably in recent years (Delen & Bellibaş, 2015; Double et al., 2020; Lee et al., 2020; Ozan & Kınca, 2018). In parallel to the publication of more research studies, a meta-analysis research study was developed in the present study. In this regard, the purpose of the study was to provide a synthesis of the experimental and quasi-experimental studies about the effectiveness of formative assessment practices on student learning in Turkey. In addition to the effectiveness of formative assessment in each education level from primary to tertiary, features of formative assessment interventions and publication types were also examined as moderator variables in the study.

In this study, how formative assessment practices in Turkey's education system promote student learning was investigated through meta-analysis. Therefore, the present study is of high importance to gain a better understanding of the effect of formative assessment practices on student learning. Examining the effectiveness of formative assessment and its moderators (i.e. types of formative assessment interventions, education level) would contribute to the literature. In this sense, the following research questions were asked in this study:

- 1) What effect do formative assessment interventions have on student learning according to the findings of the experimental studies applied in Turkey?
- 2) Do the findings of the experimental studies applied in Turkey about the effect of formative assessment interventions on student learning differ significantly according to moderating variables (features of formative assessment interventions, education level, and publication type)?

2. METHOD

Meta-analysis method was conducted in the present study. Meta-analysis is more than a statistical technique that synthesizes a series of research studies answering the same research question in a systematic way (Borenstein et al., 2009; Glass et al., 1981). This statistical method called as quantitative research synthesis helps to summarize and compare the results of the studies. When compared with other research synthesis, meta-analysis focuses on research outcomes to draw conclusions with effect sizes (Card, 2012). ProMeta3 (professional statistical software) was used for data analysis in the present study.

Several steps were carried out to perform meta-analysis (Field & Gillett, 2010); namely, (1) doing a literature review to formulate a problem; (2) specifying inclusion/exclusion criteria; (3) calculation of effect size for each study; (4) doing meta-analysis; (5) assessing moderator variables with advanced analysis; (6) doing publication bias analysis; and (7) writing the results.

2.1. Literature Review

First, research studies that investigated the relationship between formative assessment practices and student learning were collected through a search of databases. Key words were specified in English and Turkish as “formative assessment” and “biçimlendirici değerlendirme”, and “experimental” and “deneysel”, respectively. These databases are Google Scholar, PsycINFO, Turkish Council of Higher Education (YÖK) National Thesis Center, Education Research Complete, ERIC (2020), Web of Science, ULAKBİM (2020), and EBSCO (2020). Peer-reviewed journals, master's theses, and doctoral dissertations were included in the meta-analysis.

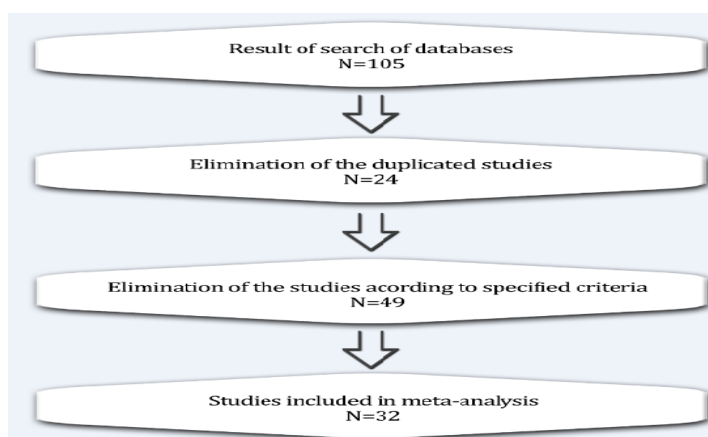
2.2. Selection Criteria

If studies had to meet the following criteria, they were included in the meta-analysis. These criteria were as follows: (1) studies that had true experimental or quasi experimental design with control group and treatment group with formative assessment interventions; (2) studies that measured learning outcomes; (3) studies with enough information to calculate effect sizes; (4) students at different education levels (i.e., primary, secondary, and tertiary); and (5) studies written in English or Turkish language.

A number of 105 records were identified through the search of databases. The number of studies dropped to 81 after removing duplicates and eliminating studies according to the specified criteria (i.e. studies that do not have formative assessment intervention, studies that do not have student learning or academic achievement, and studies that do not have enough statistics). Ultimately, 32 studies with 47 effect sizes that were unpublished theses and peer-reviewed articles that had experimental studies about the effectiveness of formative assessment on student learning were included. A flow chart that summarizes the inclusion of studies through search in the meta-analysis is given in [Figure 1](#). Therefore, the data included 32 studies as shown in [Table 1](#).

Table 1. *The studies included in the meta-analysis.*

Included Study	Number of Effect Sizes	Course	Education Level
Arıcı and Kaldırım (2015)	1	Language	Tertiary
Atik and Erkoç (2017)	2	Science	Secondary
Aydın et al.(2016)	1	Science	Secondary
Batıbay (2019)	1	Literacy	Secondary
Bayat (2014)	1	Literacy	Tertiary
Bayrak et al. (2019)	2	Science	Secondary
Baysal (2020)	1	Foreign Language	Secondary
Bolat et al. (2017)	1	Computer Science	Tertiary
Demirkesen (2019)	1	Foreign Language	Tertiary
Elvan (2012)	1	Social Sciences	Secondary
Eraz and Öksüz (2015)	1	Mathematics	Primary
Güzel (2018)	1	Science	Secondary
Hotaman (2020)	1	Teacher Training	Tertiary
Kaya and Ateş (2016)	1	Language	Primary
Kıncal and Ozan (2018)	1	Measurement and Evaluation	Tertiary
Korkmaz et al. (2019)	1	Foreign Language	Secondary
Köksalan (2019)	1	Physics	Secondary
Kuzudişli (2019)	2	Science	Secondary
Müldür and Yalçın (2019)	1	Language	Secondary
Ozan and Kıncal (2018)	1	Social Sciences	Secondary
Özgür (2016)	1	Computer Education	Tertiary
Sever and Memiş (2013)	4	Language	Primary
Tavşanlı (2019)	1	Language	Primary
Topal (2020)	1	Educational Sciences	Tertiary
Turan and Sakız (2014)	2	Science	Secondary
Yalaki and Bayram (2015)	1	Chemistry	Tertiary
Yaşar (2018)	4	Mathematics	Secondary
Yıldız and Kılıç Çakmak (2019)	1	Project Management and Application	Tertiary
Yılmaz (2015)	1	Mathematics	Secondary
Yorgancı (2015)	1	Mathematics	Tertiary
Yurdabakan and Cihanoğlu (2009)	6	Foreign Language	Secondary
Yurdabakan and Olgun (2011)	1	Science	Primary

Figure 1. Flow chart of inclusion of the studies.

2.3. Formative Assessment Interventions

Formative assessment interventions have several types of feedback. The sources of formative feedback could come from teachers, self, peers, computers, or mixed (Andrade, 2010; Graham et al., 2015). In the present meta-analysis study, the studies that have various formative feedback from adults (teachers), computers, students, and mixed are coded. The studies that have multiple interventions such as self-assessment, peer assessment, group assessment, adult feedback, and/or computer feedback were coded as mixed.

2.4. Statistical Analysis of Effect Sizes

In meta-analysis, several standardized effect sizes are used to summarize direction and magnitude of effects in research studies such as Cohen's d , Hedge's g or Glass's g (Başol-Göçmen, 2004; Lipsey & Wilson, 2001). Hedge's g also called unbiased d was used to calculate the standardized mean differences between treated groups that have formative assessment interventions and control groups. When comparing Hedge's g statistic to Cohen's d and Glass's g statistic, Hedge's g uses the pooled standard deviation (Hedge, 1981). Hedge's g is preferred since it is better for small samples (<20) and significant for different sample sizes. Hedge's g , Cohen's d , and Glass's g statistic results are interpreted in the same way. Therefore, Cohen's proposal to classify the magnitude of effects was adopted in the study (Cohen, 1987). Magnitude of effects is described as small effect (.18), medium effect (.48), and large effect (.83) in social sciences (Cohen, 1962, 1987).

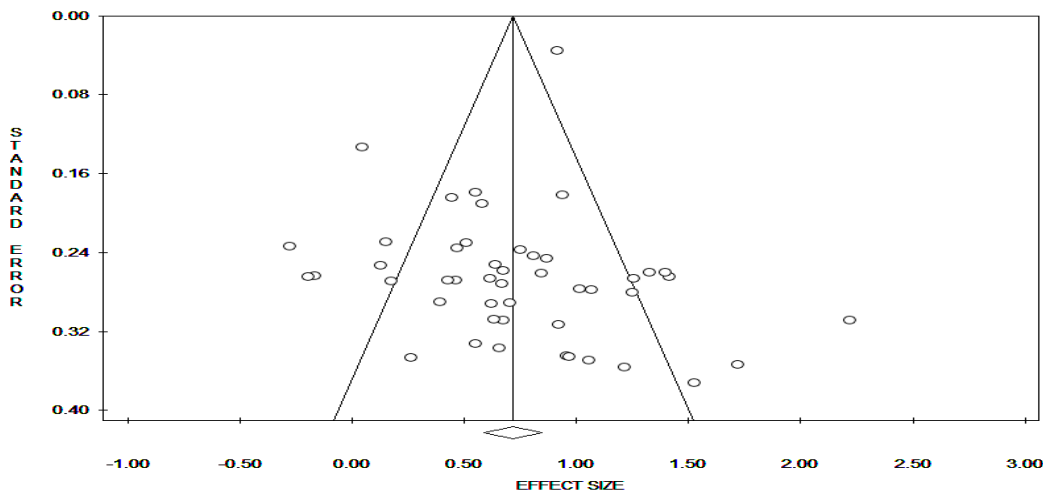
Two statistical models are used in meta-analysis. These models are fixed effect model and random effects model (Borenstein et al., 2009; Hedges & Vevea, 1998). It is assumed that only one true effect size exists for all studies including the meta-analysis with fixed effect model. On the other hand, true effect shows differences from one study to another study for the random effects model. Effect size might change due to the differences of studies such as studies that have different ages, education levels, income levels of participants, or differences of interventions. (Borenstein et al., 2009; Üstün & Eryılmaz, 2014). Due to the differences of studies, different effect sizes may occur in these studies. Therefore, estimating the mean distribution of effects is important for random effects model. Since research studies have different designs of formative assessment interventions and education levels, differences may occur from one study to another study in the present meta-analysis. For that reason, random effects model was employed in the present study. An average weighted effect size was calculated for the efficacy of formative assessment treatment. To test the heterogeneity in effect sizes, Q and I^2 statistic were used. A statistically significant p value for Q statistic means that the true effects vary (Borenstein et al., 2009). In other words, significant p value means that there is a significant variability among the effect sizes. I^2 statistic which gives the amount of

variance across studies due to heterogeneity was also computed (Higgins et al., 2003; Schwarzer et al., 2015).

2.5. Publication Bias

To have an accurate synthesis of studies in meta-analysis, assessing publication bias risk in the studies is important (Borenstein et al., 2009). There are several methods to assess the potential bias for a meta-analysis study. One of the methods is the funnel plot that gives the relationship between the observed effect size of each study and its standard error (Schwarzer et al., 2015). If studies were distributed symmetrically around the mean effect size in the plot, this would be the evidence of absence of publication bias. Funnel plot was used in the present study to inspect whether publication bias exists or not (Figure 2).

Figure 2. A funnel plot indicating standard error and observed effect size.



The funnel plot shows that studies were approximately scattered around the mean effect size. Since the interpretation of funnel plot could be subjective, some of the tests were also used to assess exactly any risks of bias such as Duval and Tweedie's Trim and Fill, Rosenthal's Fail-safe N test, Begg and Mazumdar Rank Correlation Test, and Egger's linear regression test (Begg, & Mazumdar, 1994; Duval, & Tweedie, 2000; Egger et al., 1997; Rosenthal, 1979). Trim and Fill method was used to remove extreme studies and estimate the effect sizes again in order to solve the asymmetry in funnel plot. The results of this method showed that trimming was not performed. Rosenthal's Fail-safe N test estimates how many missing studies would be needed to add to nullify the effect (Rosenthal, 1979). Rosenthal (1979) suggested that if the Fail-safe N test shows that large numbers of studies are needed to nullify the common effect rather than a few studies (i.e. five or ten), it can be concluded that true effect may not be zero in the study (Borenstein et al., 2009). In the present meta-analysis, the number of studies was 5681 according to Rosenthal's method. Therefore, it can be said that the results of the meta-analysis with 47 effect sizes would not be robust if 5681 studies were added. Besides, Egger's linear regression test was not statistically significant ($b = -0.66$, $p = 0.155$). As a result, Funnel Plot, Trim and Fill Method, Fail-safe N Test, and Egger's Linear Regression Test generally showed a low risk of publication bias that could be negligible.

3. FINDINGS

The number of studies included in the meta-analysis and the characteristics of these studies are summarized in Table 2. Most of the studies included in the meta-analysis were journal articles (68.75 %) and master's and doctoral theses (31.25 %) as publication type. 50 % of these studies were conducted at secondary school level, 34.37 % at tertiary level, and 15.62 % at primary

school level. The studies with the treatment groups having various formative assessment interventions were also described in the meta-analysis. The features of formative assessment interventions showed that 37.5 % of these studies have adult (teacher) initiated feedback, 31.25 % have computer initiated feedback, 15.62 % have student initiated feedback, and 15.62 % have mixed feedback (including peer assessment, group assessment, teacher’s feedback, and/or computer feedback).

Table 2. Characteristics of the studies included in the meta-analysis.

		Frequency (f)	Percent (%)
Study type	Thesis (master’s and doctoral)	10	31.25
	Article	22	68.75
Education level	Primary	5	15.62
	Secondary	16	50
	Tertiary	11	34.37
Features of formative assessment interventions	Adult initiated feedback	12	37.5
	Student initiated feedback	5	15.62
	Computer initiated feedback	10	31.25
	Mixed feedback	5	15.62

As summarized in Table 3, the meta-analysis shows the overall effect size as .79 with standard error of .03 in the fixed model. Heterogeneity test was used to investigate the heterogeneity in effect size. The Q value was 188.91 with 46 degrees of freedom, and p value under .05 showed heterogeneity among the studies. In other words, true effect size may have varied across studies. Besides, I^2 statistic was estimated as 75.65% indicating that the percent of variance due to between-subject factors was large. The results revealed that the impact of formative assessment on student learning varied from one study to another. By using random effects model, overall meta-analysis showed that there was a significant effect of formative assessment on student learning ($g = .72$, $SE = .07$, 95% CI = .59; 85, $p < .05$).

Table 3. Overall effect sizes and heterogeneity results related to the effectiveness of formative assessment practices.

Model	k	Mean ES	SE	Lower Limit	Upper Limit	p	Heterogeneity			
							Q value	df	p	I^2
Fixed	47	.79	.03	.74	.84	.00	188.91	46	.00	75.65
Random	47	.72	.07	.59	.85	.00				

* $p < .05$; k= number of effects; ES= effect size

In Figure 3, the forest graph demonstrating the effect size of each study based on the random effects model is presented. 3253 participants were involved in the analysis ($N_1 =$ Experimental group and $N_2 =$ Control group). It can be seen that overall effect size in random effects model across studies has a moderate level in favor of the experimental group ($g = .72$, $p < .05$).

As mentioned in Figure 3, heterogeneity test showed that the effectiveness of formative assessment practices varied from one study to the other. To investigate this variation, subgroup analysis was conducted in the present study. It is hypothesized that this variation may be explained with the studies that used various formative assessment interventions applied to different education levels and publication types. Mixed effect analysis based on random effects weights within subgroups was used to test the model. The results are presented for the features of formative assessment interventions in Table 4.

Figure 3. The forest graph showing the effect size of each study in meta-analysis.

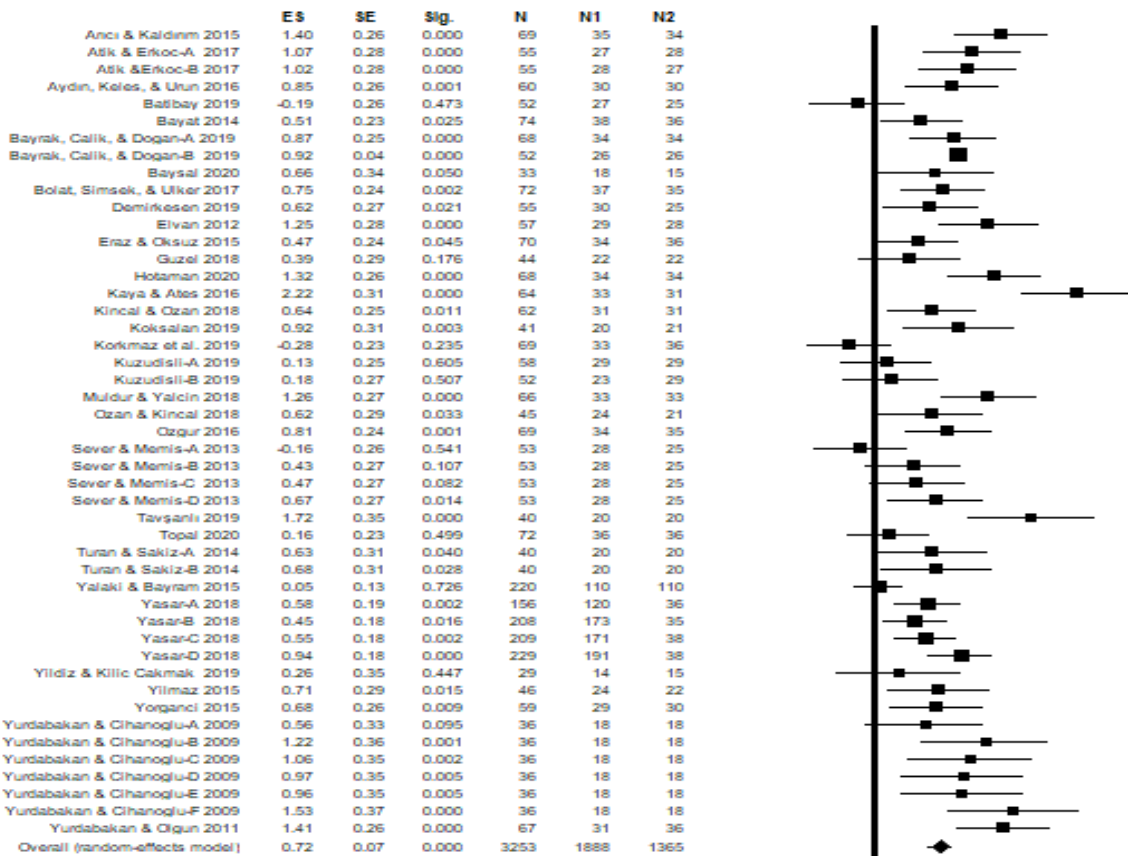


Table 4. Results of the subgroup analysis for features of formative assessment interventions.

	k	Mean ES	SE	Lower Limit	Upper Limit	p	Q value	df	p
Adult initiated feedback	20	.69	.09	.50	.87	.000			
Student initiated feedback	10	1.16	.17	.83	1.49	.000			
Computer initiated feedback	12	.42	.14	.14	.71	.003			
Mixed feedback	5	.83	.19	.46	1.21	.000			
Heterogeneity test							11.54	3	.009

*p < .05; k= number of effects; ES= effect size; SE= standard error

Mean effect size for each group was estimated by the mixed effects model. The effect sizes for each formative assessment interventions that varied between 1.16 and .42 were statistically significant. The results of the subgroup analysis might suggest that student initiated formative feedback ($d=1.16, p<.05$) and mixed feedback ($d=.83, p<.05$) had a large effect followed by a medium effect of adult feedback ($d=.69, p<.05$), and a small effect of computer feedback ($d=.42, p<.05$) on student academic performance. To compare the effect size for the subgroups, heterogeneity test was used. Total between tests ($Q=11.54, df=3, p<.05$) showed that the effect size might have varied by formative assessment intervention subgroups. In other words, features of formative assessment interventions such as adult initiated, student initiated, computer initiated, and mixed formative assessment differed significantly in the magnitude of effects.

In Table 5, mixed effects analysis was also used to estimate the effect size of groups in terms of their education levels (primary, secondary, and tertiary). The mean effect sizes that were estimated at different education levels ranged between .89 and .64 and were statistically significant. The results showed that the effect size at primary level had a large effect ($d=.89, p<.05$), while it had a medium effect on student academic performance at secondary level ($d=.71, p<.05$) and tertiary level ($d=.64, p<.05$). The results of the heterogeneity test yielded that comparison of subgroups at different education levels did not make a significant contribution to the variance ($Q=.66, df=2, p=.71$).

Table 5. Results of the subgroup analysis for education level.

Education Level	k	Mean ES	SE	%95 Confidence Interval		p	Heterogeneity		
				Lover Limit	Upper Limit		Q value	df	p
Primary	8	.89	.27	.37	1.41	.00			
Secondary	28	.71	.07	.56	.86	.00			
Tertiary	11	.64	.14	.36	.92	.00			
							.66	2	.71

* $p < .05$

The studies that included the meta-analyses were grouped in terms of publication type: articles, and theses (master’s and doctoral) (see Table 6). Mixed effect analysis showed that effect sizes according to these groups ranged between .78 and .57 and were statistically significant. The magnitude of effect size showed that articles have higher effect ($d=.78, p<.05$) than that of the theses ($d=.57, p<.05$). Heterogeneity test also showed that effect sizes of subgroups according to their publication type did not make a significant contribution to the variance ($Q=2.27, df=1, p=.13$). In other words, the distribution of effect sizes of studies according to publication type was homogeneous.

Table 6. Results of the subgroup analysis for publication type.

	k	Mean ES	SE	%95 Confidence Interval		p	Heterogeneity		
				Lover Limit	Upper Limit		Q value	df	p
Article	33	.78	.08	.63	.94	.00			
Thesis	14	.57	.11	.35	.80	.00			
							2.27	1	.13

* $p < .05$

4. DISCUSSION and CONCLUSION

In the meta-analysis study, 32 studies with a total of 47 effect sizes that met the inclusion criteria were estimated. The results of the study showed the overall mean effect size of .72. The overall mean effect size was consistent with previous meta-analysis results that effect sizes of the effectiveness of formative assessment ranged between .40 and .70 (Black & William, 1998; Graham et al., 2015). Besides, subgroup analysis was used to estimate whether the mean effect size was influenced by the features of formative assessment interventions, education level, and publication type.

The meta-analysis results showed how effective the features of formative assessment interventions were on student learning. The impact of different types of formative assessment interventions on student learning varied. The feedback from the students had the largest effect

but the feedback from the computers had the smallest effect on student learning. Moderator analysis showed that the effect sizes made a significant difference as to different types of formative feedback. Variation in effect sizes may be related to the features of formative assessment interventions in the present study. The impact of features of formative assessment interventions on student learning was examined in a few meta-analysis studies (Graham et al., 2015; Klute et al., 2017; Lee et al., 2020). Lee et al. (2020) examined various formative assessment feedback by using meta-regression. They found a similar result that the effect of student-initiated formative assessment feedback was significantly higher than teacher's formative assessment feedback and mixed feedback from both students and teachers. They implied that learner's active role is important for successful formative assessment based on their findings. Graham et al. (2015) examined four types of formative assessment feedback and found that feedback that came from teachers had the largest impact, but the feedback that came from computers had the smallest impact on student learning. In addition, by using meta-regression they also found that the effect size was not statistically related to grade level, types of formative feedback, or study quality. The present study generally shows similar results with the previous meta-analysis studies. It suggests that various formative assessment interventions in classrooms were effective. When comparing the formative assessment practices, the effect of student initiated formative interventions such as self-assessment, peer assessment, and group assessment was significantly higher than the other formative assessment interventions. Teachers, learners, and peers all have a crucial role for effective formative assessment (Black & William, 2009). The findings specifically indicated that learners' active role is very important for successful formative assessment (Clark, 2012; Lee et al., 2020).

The present study investigated that how mean effect size was in different education levels. While the highest mean effect size was found at primary school level, the lowest mean effect size was found at tertiary level. Mixed effects analysis was used to examine whether group differences were significant or not. The results showed that education level did not make a significant contribution to the variance. Likewise, King and Nash (2011) found that grade level did not make a difference on the effect of formative assessment on student learning. It can be interpreted that formative assessment is effective for all levels of education (Black & William, 1998; King & Nash, 2011). Therefore, the number of using formative assessment in classrooms should be increased in all levels of education.

Lastly, the studies included in the meta-analysis were categorized into two groups (published articles versus theses). Heterogeneity test showed that effect sizes of studies according to their publication type were homogeneous. The magnitude of the effect size did not significantly differ between the published articles and unpublished theses. It can be concluded that this finding was resistant to file drawer treat (publication bias) (Rosenthal, 1979).

Briefly, the present meta-analysis synthesized research studies conducted in Turkey showed that formative assessment interventions have a positive impact on student learning for all education levels. Assessment for learning rather than assessment of learning has been more emphasized in Turkey's curriculum since the 2005 educational reform. Assessment for learning strategies that curriculum requires has become widespread from primary schools to universities (MoNE 2013, 2017, 2018, 2020, YÖKAK 2018, 2019). The present study could give evidence regarding the effectiveness of formative assessment on student learning in Turkey's education system. Besides, there are only a few studies that examined the effectiveness of formative assessment interventions types. Therefore, it suggests that more meta-analysis studies should be conducted on this area (Lee et al., 2020). The present meta-analysis study is promising to provide a significant contribution to literature regarding the effectiveness of formative assessment interventions types. That is why more empirical studies are needed to have evidence regarding the effectiveness of formative assessment practices. Increasing different types of

formative assessment practices especially encouraging learners to have an active role in this process (i.e. self-assessment, peer-assessment, and group-assessment) is crucial. Since the results suggest that use of formative assessment strategies is effective for all education levels, implementation of formative assessment activities efficiently in classrooms is also important. Thus, providing all teachers and college scholars with professional development as to how to use formative assessment tools is highly needed.

The findings of this meta-analysis study were limited by the number of studies on formative assessment conducted in Turkey. Another limitation in the study was examining a few moderator variables such as education level, types of formative assessment interventions, and publication types. In further meta-analysis research, investigation and comparison of more variables such as subject areas, formality of formative assessment, feedback procedures, and feedback time are needed.

Declaration of Conflicting Interests and Ethics

The author declares no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author.

ORCID

Pinar Karaman  <https://orcid.org/0000-0002-2218-2701>

5. REFERENCES

- Andersson, C., & Palm, T. (2017a). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and Instruction, 49*, 92-102. <https://doi.org/10.1016/j.learninstruc.2016.12.006>
- Andersson, C. & Palm, T. (2017b). Characteristics of improved formative assessment practice. *Education Inquiry, 8* (2), 104-122. <https://doi.org/10.1080/20004508.2016.1275185>
- Andrade, H. (2010). Students as the definite source of formative assessment: Academic self-assessment and the self-regulation of learning. In G. J. Cizek & H. L. Andrade (Eds.), *Handbook of formative assessment* (pp. 90–105). Routledge Publishing.
- Andrade, H., & Brookhart, S. M. (2016). The role of classroom assessment in supporting self-regulated learning. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 293–309). Springer Publishing.
- Başol-Göçmen, G. (2004). Meta-analizinin genel bir değerlendirmesi [A general revision of meta-analysis]. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi, 7*, 186–192.
- Begg, C.B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101. <https://doi.org/10.2307/2533446>
- Bennett, R. E. (2011). *Formative assessment: a critical review. Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25. <https://doi.org/10.1080/0969594X.2010.513678>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. <https://doi.org/10.1007/s11092-008-9068-5>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley and Sons Publishing.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: brain, mind, experience, and school (expanded ed.)*. National Academy Press.
- Brookhart, S. M. (2009). *Exploring formative assessment*. ASCD Publishers.

- Brookhart, S.M. (2010). *Formative assessment strategies for every classroom (2nd ed.)*. An ASCD Action Tool Publishers.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245-281. <https://doi.org/10.3102/00346543065003245>
- Card, N. A. (2012). *Applied meta-analysis for social science research*. The Guilford Press.
- Cauley, K.M. & McMillan, J.H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 83 (1), 1-6. <https://doi.org/10.1080/00098650903267784>
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205–249.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *Journal of Abnormal and Social Psychology*, 65, 145–153.
- Cohen, J. (1987). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates.
- Delen, I., & Bellibas, M. S. (2015). Formative assessment, teacher-directed instruction and teacher support in Turkey: Evidence from PISA 2012. *Mevlana International Journal of Education*, 5(1), 88-102. <http://dx.doi.org/10.13054/mije.15.01.5.1>
- Double, K., McGrane, J. & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32 (2), 481–509. <https://doi.org/10.1007/s10648-019-09510-3>
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109), 629-634.
- Field, A.P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, 63, 665–694. <https://doi.org/10.1348/000711010X502733>
- Fuchs, L., & Fuchs, D. (1986). Effects of systematic formative evaluation on student achievement: A meta-analysis. *Exceptional Children*, 53, 199-208. <https://doi.org/10.1177/001440298605300301>
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20, 304-315. <https://doi.org/10.1016/j.learninstruc.2009.08.007>
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Sage Publications.
- Graham, S., Hebert, M., & Harris, K.R. (2015). Formative assessment and writing. *The elementary school journal*, 115 (4), 523-547.
- Hargreaves, E. (2008). Assessment. In G. McCulloch, & D. Crook (Eds.), *The routledge international encyclopedia of education (pp. 37–38)*. Routledge Publishing.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 metaanalyses relating to achievement*. Routledge Publishing.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hedges, L.V. (1981). Distribution theory for Glass' estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486-504. <https://doi.org/10.1037/1082-989X.3.4.486>
- Higgins, J., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327, 557–560. <https://doi.org/10.1136/bmj.327.7414.557>

- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28-37. <https://doi.org/10.1111/j.1745-3992.2011.00220.x>
- Kitchen, H., Bethell, G., Fordham, E., Henderson, K., & Li, R.R. (2019). *OECD reviews of evaluation and assessment in education: student assessment in Turkey*, OECD reviews of evaluation and assessment in Education, OECD Publishing. Retrieved August 17, 2021 from <https://doi.org/10.1787/5edc0abe-en>
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>
- Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017). Formative assessment and elementary school student academic achievement: A review of the evidence (REL 2017–259). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved January 15, 2021 from <https://files.eric.ed.gov/fulltext/ED572929.pdf>
- Lee, H., Chung, H.Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: systematic review. *Applied Measurement in Education*, 33(2), 124-140. <https://doi.org/10.1080/08957347.2020.1732383>
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Sage Publications.
- Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education*, 95, 85–98. <https://doi.org/10.1016/j.compedu.2015.12.002>
- McManus, S., Ed. (2008). *Attributes of effective formative assessment*. Council of Chief State School Officers.
- Miller, T. (2009). Formative computer-based assessment in higher education: the effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education*, 34 (2), 181-192. <https://doi.org/10.1080/02602930801956075>
- MoNE (2013). Early childhood education program [Okul öncesi eğitim programı], Ministry of National Education, Ankara. <https://tegm.meb.gov.tr/dosya/okuloncesi/ooproram.pdf>
- MoNE (2017). The topics of in-service training activities in the last five years, 2012-2016 [Son 5 yılda düzenlenen (2012-2016) hizmetiçi eğitim faaliyetleri konuları], Ministry of National Education, Ankara.
- MoNE (2018). Geography curriculum [Coğrafya dersi öğretim programı], Ministry of National Education, Ankara. <https://mufredat.meb.gov.tr/Dosyalar/2018120203724482-Cografya%20dop%20pdf.pdf>
- MoNE (2020). Strengthening teacher capacity based on school and classroom-based assessment. Social Studies Lesson Teacher's Guide Booklet [Okul ve sınıf tabanlı değerlendirmeye dayalı öğretmen kapasitesinin güçlendirilmesi. Sosyal Bilgiler Dersi Öğretmen Rehber Kitapçığı], Ministry of National Education, Ankara
- Ozan, C., & Kıncal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory & Practice*, 18, 85–118. <http://dx.doi.org/10.12738/estp.2018.1.0216>
- Popham, W. J. (2008). *Transformative assessment*. Association of Supervision and Curriculum Development.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results, *Psychological Bulletin*, 86(3), 638-641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18 (2), 119–144. <http://dx.doi.org/10.1007/BF00117714>

- Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. Cham: Springer.
- Üstün, U., & Eryılmaz, A. (2014). A research methodology to conduct effective research syntheses. *Education and Science*, 39(174), 1-32.
- Weldmeskel, F.M., & Michael, D.J. (2016). The impact of formative assessment on self-regulating learning in university classrooms. *Tuning Journal for Higher Education*, 4 (1), 99-118. [https://doi.org/10.18543/tjhe-4\(1\)-2016pp99-118](https://doi.org/10.18543/tjhe-4(1)-2016pp99-118)
- Wiliam, D. (2018). Feedback: at the heart of –but definitely not all of–formative assessment. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 3–28). Cambridge University Press.
- Van der Kleij, F., Feskens, R., & Eggen, T.J.H.M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research* 85(4), 1-37. <https://doi.org/10.3102/0034654314564881>
- YÖKAK (2018). Yükseköğretim değerlendirme ve kalite güvencesi 2017 yılı durum raporu [2017 higher education evaluation and quality Assurance Status Report], Ankara: YÖK. https://yokak.gov.tr/Common/Docs/Site_Activity_Reports/2018DurumRaporuv2.pdf
- YÖKAK (2019). Yükseköğretim değerlendirme ve kalite güvencesi 2018 yılı durum raporu [2018 Higher Education Evaluation and Quality Assurance Status Report], Ankara: YÖK.
- Zimmerman, B. (2002). Becoming a self-regulated student: An overview. *Theory into Practice* 41(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2
- Zimmerman, B., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, 31(4), 845-862. <https://doi.org/10.3102/00028312031004845>

6. APPENDIX

Studies included in the meta-analysis marked with an *

- *Arici, A.F., & Kaldirim, A. (2015). The effect of the process-based writing approach on writing success and anxiety of pre-service teachers. *Anthropologist*, 22(2), 318-327. <https://doi.org/10.1080/09720073.2015.11891883>
- *Atik, A.D., & Erkoç, F. (2017). The impact of formative tests on student achievement. *Journal of Theory and Practice in Education*, 13(4), 670-692.
- *Ayдын, S., Ural Keleş, P., & Ürün, N. (2016). Süreç değerlendirme yönteminin 7. Sınıf öğrencilerin güneş sistemi ve ötesi: uzay bilmecesi ünitesinde akademik başarıları ve kalıcılık düzeylerine etkisi [The effect of formative assessment on the achievement and retention levels of 7th grade students at the unit of solar system and beyond: mystery in space]. *Türk Eğitim Araştırmaları Dergisi (TURKEAD)*, 1(1), 11-17.
- *Batıbay, E.F. (2019). *Web 2.0 Uygulamalarının Türkçe dersinde motivasyona ve başarıya etkisi: kahoot örneği [The impact of Web 2.0 applications on motivation and success in Turkish course: the example of kahoot]* [Master's thesis]. Hacettepe University.
- *Bayat, N. (2014). The effect of the process writing approach on writing success and anxiety. *Educational Sciences: Theory & Practice*, 14(3), 1123-1141.
- *Bayrak, N., Çalık, M., & Doğan, S. (2019). The effects of smart formative assessment system on academic achievement and course process. *Hacettepe University Journal of Education*. Advance online publication. <https://doi.org/10.16986/HUJE.2019056742>.
- *Baysal, H. (2020). *Altıncı sınıf İngilizce dersinde kavram karikatürleri kullanımının öğrenci başarısına, konuşma becerisine ve motivasyonuna etkisi [The effect of using concept cartoons on students' achievement, speaking skill, and motivation in the sixth grade English]* [Master's thesis]. Balıkesir University.
- *Bolat, Y.İ., Şimşek, Ö., Ülker, Ü. (2017). Oyunlaştırılmış çevrimiçi sınıf yanıtı sistemine akademik başarıya etkisi ve sisteme yönelik görüşler [The impact of gamified online classroom response system on academic achievement and views about this system]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 17(4), 1741-1761.
- *Demirkese, B. (2019). *The effects of a mobile phone application on Turkish EFL students' grammar learning* [Unpublished master's thesis]. Necmettin Erbakan University.
- *Elvan, Ö. (2012). *Sosyal Bilgiler öğretiminde çalışma yaprakları kullanılmasının kavram yanlışlarını gidermeye etkisi [The effect of the usage of worksheets for resolving misconceptions in teaching social studies]* [Master's thesis]. Ahi Evran University.
- *Eraz, G., & Öksüz, C. (2015). Effect of primary school teachers' feedback on students' extracurricular mathematics activities. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 36, 105-119.
- *Güzel, Z. (2018). *Fen bilimleri öğretiminde öz ve akran değerlendirme uygulamalarının yer aldığı probleme dayalı öğrenme yaklaşımının öğrencilerin başarı ve tutumlarına etkisi [The effects of problem based approach practiced through self and peer assessment on student achievement and attitudes in science teaching]* [Master's thesis]. Necmettin Erbakan University.
- *Hotaman, D. (2020). The effect of formative assessment on the academic achievement levels of prospective teachers. *Journal of Curriculum and Teaching*, 9(3), 33-44.
- *Kaya, B., & Ateş, S. (2016). The effect of process-based writing focused on metacognitive skills oriented to fourth grade students' narrative writing skill. *Education and Science*, 41(187), 137-164. <https://doi.org/10.15390/EB.2016.6752>

- *Kıncal, R.Y. & Ozan, C. (2018). Effects of formative assessment on prospective teachers' achievement, attitude and self-regulation skills. *International Journal of Progressive Education*, 14(2), 77-92. <https://doi.org/10.29329/ijpe.2018.139.6>
- *Korkmaz, Ö., Vergili, M., Çakır, R., & Uğur Erdoğan, F. (2019). Plickers Web 2.0 ölçme ve değerlendirme uygulamasının öğrencilerin sınav kaygıları ve başarıları üzerine etkisi [The impact of plickers Web 2.0 assessment and evaluation tool on exam anxiety and academic success of students]. *Gazi Eğitim Bilimleri Dergisi*, 5(2), 15-37. <https://dx.doi.org/10.30855/gjes.2019.05.02.002>.
- *Köksalan, S. (2019). *Sorgulamaya dayalı öğretimde kullanılan biçimlendirici değerlendirmenin öğrencilerin Fizik dersine yönelik tutumlarına ve kavramsal öğrenmelerine etkisinin incelenmesi* [Investigation of the effect of formative assessment used in inquiry-based instruction on students' attitudes towards physics lesson and conceptual learning] [Master's thesis]. Marmara University.
- *Kuzudişli, H. (2019). *Video-içi biçimlendirici değerlendirme ortamında öğrenen değerlendirme etkileşimlerinin incelenmesi* [Investigating of interaction between learner-assessment in the video formative assessment environment] [Master's thesis]. Hacettepe University.
- *Müldür, M., & Yalçın, A. (2019). Öz düzenlemeye dayalı yazma eğitiminin ortaokul öğrencilerinin bilgilendirici metin yazma becerisine, yazmaya yönelik öz düzenleme becerisine ve yazma öz yeterlik algısına etkisi [The effect of self-regulated writing instruction on middle school students' informative writing skills, self-regulated writing skills, and self-efficacy perception]. *Ilkogretim Online*, 18(4), 1779-1804. <https://dx.doi.org/10.17051/ilkonline.2019.639323>
- *Ozan, C., & Kıncal, R. Y. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences: Theory & Practice*, 18, 85–118. <http://dx.doi.org/10.12738/estp.2018.1.0216>
- *Özgür, P. (2016). Facebook sosyal ağına entegre e-portfolyo yazılımının akademik başarı ve öğretim sürecinde kullanımına yönelik tutuma etkisi [The effect of e-portfolyo software integrated to facebook social network on academic success and attitudes towards its use in teaching process]. *Sakarya University Journal of Education*, 6(1), 38-56.
- *Sever, E., & Memiş, A. (2013). Süreç temelli yazma modellerinin ilkökul dördüncü sınıf öğrencilerinin yazım–noktalama becerisine ve yazma eğilimine etkisi [The Effects Of Process-Based Writing Models On Primary School 4th Grade Students' Spelling-Punctuation Skills And Writing Dispositions]. *Karadeniz Sosyal Bilimler Dergisi*, 5(9), 243- 259.
- *Tavşanlı, F. (2019). *Süreç temelli yazma modüler programının ilkökul 2. sınıf öğrencilerinin yazmaya ilişkin tutum, yazılı anlatım becerisi ve yazar kimliği üzerine etkisi* [The effect of process writing modular program on 2nd grade elementary school students' towards attitudes, writing skills and their author identity] [Doctoral dissertation]. Uludağ University.
- *Topal, M. (2020). *Oyunlaştırma ile zenginleştirilmiş çevrimiçi öğrenmenin başarı, çevrimiçi bağlılık ve öğrenme motivasyonu üzerinde etkisi* [The effect of online learning enhanced with gamification on student's engagement to online learning environment, academic achievement and learning motivation] [Doctoral dissertation]. Sakarya University.
- *Turan, M.A., & Sakız, G. (2014). Fen ve teknoloji dersinde portfolyo kullanımının öğrenci başarısı ve kalıcılığa etkisi [The influence of portfolios on student success and retention level in science and technology class]. *Mersin University Journal of the Faculty of Education*, 10(3), 48-63.

- *Yalaki, Y. & Bayram, Z. (2015). Effect of formative quizzes on teacher candidates' learning in general chemistry. *International Journal of Research in Education and Science (IJRES)*, 1(2), 151-156.
- *Yaşar, C. (2018). *Geri bildirim verilme zamanının matematik başarısına etkisi [The effect of feedback time on mathematics achievement]* [Master's thesis]. Hasan Kalyoncu University.
- *Yıldız, G., & Kılıç Çakmak, E. (2019). Zenginleştirilmiş e-değerlendirme sisteminin ders başarısına etkisi ve öğrenci memnuniyetinin incelenmesi [The effect of enriched e-assessment system on course success and review of student satisfaction]. *Gazi Journal of Education Sciences (GJES)*, 5, 106-139.
- *Yılmaz, N. (2015). Cebir öğretiminde yazma etkinliklerini kullanmanın ortaokul 7. sınıf öğrencilerinin başarılarına etkisi [The impact of using writing activities in teaching algebra on seventh grade middle school students' achievement]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 15(1), 356-376.
- *Yorgancı, S. (2015). Web tabanlı uzaktan eğitim yönteminin öğrencilerin matematik başarılarına etkileri [The effects of web based distance education method on students' mathematics achievements]. *Kastamonu Eğitim Dergisi*, 23(3), 1401-1420.
- *Yurdabakan, İ., & Cihanoğlu, M. O. (2009). Öz akran değerlendirmenin uygulandığı işbirlikli okuma ve kompozisyon tekniğinin başarı, tutum ve strateji kullanım düzeylerine etkisi. [The effects of cooperative reading composition technique with the applications of self and peer assessment on the levels of achievement, attitude, strategy use]. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 11(4), 105-123.
- *Yurdabakan, İ., & Olgun, M. (2011). The influence of peer and self-assessment on learning and metacognitive knowledge: Consequential validity. *International Journal on New Trends in Education and Their Implications*, 2(4), 44-57.