

The Effect of Systematic Implementation of Formative Assessment on Male and Female EFL Learners' Academic Achievement

Amir Asadifard

Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran
Email: a.asadifard@gmail.com

*Akbar Afghari**

Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran
Email: afghary@yahoo.com

Abstract

Substantial claims have been made that Formative Assessment (FA), if practiced effectively, improves learning. This study aimed at exploring the effect of FA on EFL learners' academic achievement. The population included high school teachers and their students in Khorramabad, Lorestan, Iran. The sample included 40 teachers and their students (n=651) who were randomly selected and assigned to two groups of experimental and control. Teachers in the experimental group participated in a workshop on FA before the treatment. However, teachers in the control group followed the usual syllabus based on Iran's educational system. The students' academic achievement was measured in both groups by an achievement test before the treatment. Trained observers attended the classes in the experimental group and completed Formative Assessment Observation Report, which reports teachers' practice of FA. After the treatment, posttests on students' achievement in both groups were conducted. To analyze the data, ANCOVA was used. The results indicated that effective implementation of FA enhances students' academic achievement. However, gender was not a determining factor. Moreover, among FA components, *clarifying learning targets* was the most and *monitoring* was the least frequent strategy used by teachers.

Keywords: Achievement, EFL learners, Formative Assessment, Gender

1. Introduction

Understanding and implementing effective classroom assessments are essential to increasing student achievement, and teachers' assessment practices are essential elements for addressing students' learning needs. Knowing about teachers' assessment practices serves as a way of finding out if teachers adopt or use effective assessment methods and strategies in order to meet the learning needs of learners (McMillan, 2010).

One type of assessment, which seems to be effective in teaching and learning environment, is formative assessment (FA). Language practitioners, districts and school personnel are increasingly being encouraged to implement FA (Black & Wiliam, 1998a) because research has indicated that this kind of assessment is a powerful means to improve both teaching and learning situation in class. FA is supposed to be one of the effective means to boost the quality of learning and teaching and one of the effective factors in pedagogy (Oswalt, 2013; Popham, 2008). The idea of FA has been around for long; however, the use of it as a technical term is recent (McDowell, Wakelin, Montgomery, & King, 2011).

Different definitions have been presented by leading authorities such as Black and Wiliam (1998a, 1998b), and Popham (2008), as to what assessment in general and FA specifically is. For instance, Black and Wiliam (1998a) defined FA as a number of activities the teachers and their students conduct to provide information which is in turn used as feedback to modify the teaching and learning environment in a positive way. From one perspective, FA can be defined in terms of what it is not. In this sense, one of the characteristics of FA is that the evaluation aspect is absent in this kind of assessment (Cizek, 2010). In other words, penalties and grades consequences are not part of FA. FA, however, is much more than this. In a broad sense, FA is considered as a collaborative process in which educators and students are involved in order to understand the students' learning and conceptual organization, diagnosis of weaknesses, areas for improvement, identification of strengths, and as a source of information that teachers can utilize in their instructional planning. Students can also use it to deepen their understanding and improve their achievement (Cizek, 2010).

In order to define the construct of FA, operationally, we drew upon one of the most agreed-upon description. Many experts in the field believe that the following strategies should be considered and implemented by classroom teachers who claim to practice FA in

their teaching practice (Bennett, 2011). These five key strategies are:

1. Clarifying and sharing learning intentions and criteria for success.
2. Engineering effective classroom discussions, questions, and learning tasks.
3. Providing feedback that moves learners forward.
4. Activating students as the owners of their own learning.
5. Activating students as instructional resources for one another.

As reported in the literature, FA and its effects on different variables have been investigated using a variety of methods, however, research on FA and student achievement is still scant. Therefore, the present study investigated the effect of FA on male and female Iranian high school students' academic achievement.

2. Literature Review

Different theories have been proposed to be in association with FA. According to McMillan (2010), two important skills are involved when FA is practiced for deep understanding: metacognition and self-regulation. Self-regulation is a broader construct, which encompasses metacognition, self-evaluation, self-assessment, and self-reaction. Self-regulation helps students to believe that they can manage their own learning and they are equipped with needed skills to monitor and judge their performance so that they will be able to boost deep understanding (Hattie & Timperley, 2007). Deep understanding metacognition emphasizes the process not the product of learning. Therefore, students will be encouraged to develop self-appraisal and self-management skills. This, in turn, improves their self-directed learning. Students learn, "how and when to request feedback, and they become adept at error detection and corrective skills" (McMillan, 2010, pp. 47-8).

Some researchers such as Crooks (1988) reported that grading, as it is used by many teachers, may not help learners improve their knowledge. However, grades should be used to assist students to learn. Therefore, "A classroom culture in which teachers and students are partners in learning should be established" (McManus, 2008, p. 5).

On the other hand, teachers have been always in search of ways to improve student achievement as one of the most important educational goals. The relationship between FA and student achievement has been investigated from different perspectives. For instance, Davidheiser (2013) investigated assessment literacy and the role of an effective classroom assessment implementation on student achievement. In other words, it is believed that

teachers' assessment literacy is crucially important in building the connection between assessment quality and student achievement. It was suggested that creating sufficient time for teachers to learn, implement, and evaluate techniques connected to assessment is extremely important. It was also suggested that providing professional development on FA could help teachers learn new techniques and utilize them in their classes. Finally, it was emphasized that adequate levels of organizational support and leadership must be present and apparent with regard to reforming teachers' assessment practices (DeNome, 2015).

FA results seem to be a predictor for student achievement. As Smith (2013) reported, academic achievement can be accurately predicted by comprehensive post FAs. Based on the findings of the same study, grades were improved when FA strategies were used. Likewise, unit gains on post assessments demonstrated a statistically significant indicator for academic achievement on high stake standardized assessments.

Klute, Apthorp, Harlacher, Reale, and Research (2017) did a meta-analysis on the effectiveness of FA on achievement. The study focused, specifically, on elementary school student achievement followed by effective implementation of FA. Results of this study indicated that FA had a positive and significant effect on student academic achievement. It was also found that FA in math had larger effects, on average, on student academic achievement than did FA in reading and writing. Two studies investigated FA in spelling with special education students. Four studies examined FA in composition with older elementary school students in grades 4-6. Furthermore, it turned out that, among all subject areas, FA had larger effects on academic outcomes when other agents directed the FA. As for the subject areas, both student-directed and other-directed FA in math were effective. However, in reading, other-directed FA was more effective than student-directed FA. In writing, on the other hand, other-directed FA did not have substantively important effects, and not enough evidence was available to determine the effectiveness of student-directed FA.

Macgregor, Spiers, and Taylor (2011) conducted a study in which they investigated the effect of formative audio feedback on student learning. They focused on the delivery of voice emails to undergraduate students and evaluated the efficacy of such feedback in FA. The results indicated that audio feedback, which better conforms to existing models of quality formative feedback, can enhance the student learning experience and can be more efficient in feedback delivery.

DeNome (2015) reviewed the impact of professional development based on the principles of FA on student achievement. The study compared mathematics and reading performance data from student populations with teachers who were trained in FA to performance data from student populations with teachers who had not received training in FA. The study found that students who received instruction from teachers trained in FA had higher scores. The assessment data indicated a statistically significant difference in growth in the area of mathematics and reading.

Bangert-Drowns, Kulik, and Morgan (1991) strongly supported the idea that student academic achievement directly correlated to the number of FAs given. However, as they believe, the most important factor is that focus must not be on grades or scores for students comparison; the focus must be on the quality of work or concept to be mastered (Butler, 1987). According to Fisher and Frye (2007), “FAs are ongoing assessments, reviews, and observations in a classroom” (p. 4) which provide a “systematic process to continuously gather evidence and provide feedback about learning while instruction is under way” (Heritage, Kim, Vendlinski, & Herman, 2009, p. 24). Therefore, the learners should be engaged in the process. Furthermore, they should be continuously checked for understanding to see if FA has been conducted effectively.

Many subsequent studies investigated specific aspects of FA techniques and their academic learning benefits. Wiliam (2008) reported that teachers given supports to implement FA techniques were able to improve student achievement gaps by 50 percent. Another study conducted by Dunn and Mulvenon (2009), though, challenged research claims of some of the more seminal studies, most specifically *Inside the Black Box* by Black and Wiliam (1998b). Dunn and Mulvenon (2009) claimed that the findings from the two seminal works did not prove that levels of academic achievement were possible through the use of FA processes.

As reported above, a large body of research has been conducted on the inter-relationship between FA and achievement. However, research on the effectiveness of FA on language is not sufficient. In other words, language skills have been investigated separately, but the overall language achievement, which is the focus of this study, has not been fully explored. Furthermore, in most studies, the judgment of whether the teachers apply FA has been based on a single observation or an administration of a questionnaire. This way, that is, a single observation or report, we cannot be sure that the teachers have

used FA as a part of the class instruction. To alleviate this problem, in this study, a number of observations with sufficient time interval were used to make sure that FA is conducted as an integral part of the learning and teaching process. Considering the problems mentioned, in this study, attempts were made to explore if systematic and ongoing implementation of FA followed by a workshop on FA strategies can improve junior high school students' academic achievement. Additionally, the role of FA components and gender in improving FA were investigated. Therefore, based on the objectives of the study, the following research questions were formulated:

1. Does formative assessment implementation have a significant role in Iranian EFL learners' academic achievement?
2. Are there any differences between FA components in improving students' academic achievement?
3. Does gender have a role in the effect of FA on student academic achievement?

3. Methodology

3.1. Design and Context of Study

In this study, academic achievement was the dependent variable and FA was the independent variable. Both the control and the experimental students' academic achievement were measured using a pretest and posttest administration of the achievement tests. Therefore, this research was a quasi-experimental study using a pretest, treatment, and delayed posttest design carried out over a period of 12 weeks with participants who were randomly assigned to experimental and control groups.

3.2. Participants

The population of the study included all high school teachers and their students in Khorramabad, Lorestan, Iran. There were 113 high schools in the city. Of these, 57 were male-student schools and 56 were female-student schools. Ninety-two English teachers were teaching in the two districts of Khorramabad at the time of the experiment. There were 37 male and 55 female teachers. The number of the students was 5984. Twenty teachers and their students participated in the study. The number of students was 651. According to a self-report checklist filled out by these teachers, they had not been using FA systematically in their classes before this study. This sample consisted of teachers who

were teaching in 3rd grade classes of junior high schools during the experiment in 2016-2017 academic year. Identification of participating school district, teachers, and students were kept confidential. The experimental group teachers consisted of 8 male teachers and 12 female teachers. The teachers varied in years of experience (7 to 25 years of educational teaching experience).

According to Statistics Office of Lorestan Education Organization, there were 103 junior high schools in Khorramabad. Of these, 57 were male-student schools and 56 were female-student schools. Ninety-two English teachers were teaching in the two district of Khorramabad at the time of the experiment. There were 37 male teachers who were teaching in males' classes and 55 female teachers who were teaching in females' classes. The number of the students was 5984. There were 3139 male and 2845 female students.

After the teachers were asked to cooperate, through an official letter by the Education Organization of Lorestan, 40 teachers agreed to participate in this study. According to a self-report checklist filled out by these teachers, they had not been using FA systematically in their classes before the study. The teachers, then, were randomly assigned to control and experimental groups. Therefore, sampling started with availability of subjects (teachers who were willing to participate). Then, random assignments were done for the experiment and control groups.

The experimental group consisted of 8 male teachers and 12 female teachers. The average years of experience was 15.5 and the median was 13.5. The total number of student participants in the experimental group was 371. Of these participants, there were 150 male and 221 female students. It was expected that the participating teachers would systematically include the attributes of FA in their daily classroom instruction. Teachers were trained and encouraged to consider and implement the components of FA that is, articulating learning targets, monitoring, providing feedback, and promoting self- and peer-assessment.

The control group consisted of 9 male and 11 female teachers. The average years of experience for the control group was 18.7 and the median was 21 years. The total number of student participants in the control group was 280. Of this number, there were 111 male and 169 female students. Since the teachers in the control group did not take part in the workshop, they did not focus on FA strategies. However, they continued to instruct their students using the national curriculum and the country learning standards.

3.3. Instruments

Two instruments were used in this study: a) Formative Assessment Observation Report developed by Oswald (2013), and b) the achievement tests developed by the researchers. In order to collect data on the teachers' use of FA, Formative Assessment Observation Report was used. The instrument was piloted, developed, validated, and used by Oswald (2013). Oswald (2013) reports that the instrument was constructed around five components of formative assessment: understanding learning targets, monitoring student learning, feedback, self-assessment, and peer assessment. There were 20 items in the instrument rated on a 1-5 Likert-type scale. The instrument was evaluated on the basis of reliability across time, reliability across raters, and reliability of scale. In order to gather evidence regarding content validity, Oswald (2013) gave the suggested 28 items, grouped into five formative assessment components, to faculty members who were considered experts in the field. They reviewed and approved of the suggested potential items and their grouping, providing a basis for content validity for the instrument.

Two achievement tests were used to collect data on students' progress during a semester. In order to know if systematic and ongoing implementation of FA can enhance students' academic achievement, two parallel achievement tests were used prior to and after the treatment. The tests, which were parallel, included language skills and subskills covered in the course book. To measure students' academic achievement, these two tests were given to the students in the experimental as well as the control groups. The tests contained two subtests: reading and writing section, and listening section. These two subtests accounted for 60% and 40% of the total mark, respectively. Both tests contained a variety of questions which are usually used to measure students' achievement on the materials covered during the first semester of a typical academic year. The tests were based on the books which are presently taught at junior high schools across the country. These tests scores range from zero to twenty points. The tests were in compliance with those developed by the Curriculum and Textbooks Development Office of Ministry of Education. They were also given to 3 evaluators to check their congruence with the content of the book hence their validity was ensured. In addition to this, based on the scores obtained, the reliability of the tests was calculated for both tests using Cronbach alpha reliability test. The reliability of the two tests were .77 and .81, respectively.

3.4. Data Collection Procedure

Before the treatment, one of the achievement tests was administered as the pretest to the students in both experimental and control groups. During the semester, the classes in the experimental group were observed by trained observers three times. They were TEFL senior students who were studying at Tabatabaee campus of Farhangian University in Lorestan. There was a three-week time interval between observations to make sure that the teachers were utilizing FA continuously. The average score for all observations was considered as the teachers' score for FA implementation. The score was the result of the Formative Assessment Observation Report (Oswalt, 2013), which was completed by the observers in the class time.

Teachers in the experimental group utilized FA strategies based on the training they had in the workshop which was held prior to the treatment. In that workshop, they became familiar with the nature of FA and its strategies. The teacher participants were also taught how to use FA strategies in their classes. These teachers received professional development about the attributes of FA, the strategies associated with FA, and how to implement these in classrooms as part of the teaching process. Based on the reports by the class observers, the teachers in the experimental group utilized FA strategies which were presented at the workshop. Detailed information on the use of FA strategies by the teachers in the experimental group appears in Table 4 below.

The training workshop consisted of an overview of the philosophy and more specific training on the successful implementation of FA based on the works of Black and Wiliam (1998b), Heritage (2010), McManus (2008), Popham (2008), and Stiggins (2005). The first meeting occurred prior to the beginning of the school year. The experimental group teachers were presented with an overview of FA and its attributes. The workshop was scheduled to be held one week later. The workshop which lasted for 12 hours covered topics related to FA. These topics included FA and related research, the impact of FA on students' learning, the goals of implementing FA and how students can be trained to be responsible for their own learning, FA procedures and application, and instructional adjustments in doing FA.

Participating teachers in the control group continued to instruct their students as they normally did. These teachers continued to follow the school and district's curriculum guides and national education organization standards for learning. They had not received additional professional development on FA. Participating students in the control group took

part in class as normal. At the end of the semester, the posttest was conducted on students' achievement in both groups. The results of the pretests and posttests on academic achievement in the control and experimental groups were compared to discover if the systematic use of FA strategies had a statistically significant impact on students' academic achievement.

3.5. Data Analysis Procedure

To analyze the data obtained in the study, analysis of covariance (ANCOVA) was used. Pallant (2011) states that analysis of covariance is an extension of ANOVA that allows you to explore differences between group means while statistically controlling for an additional variable (in this case pretest). The scores on the pretest are treated as a covariate to control for pre-existing differences between the groups. It is also useful when it is impossible to randomly assign participants to the different groups, but instead have to use existing groups (e.g., classes of students). In the case of this study, there were differences between the scores of students' pretests hence doing ANCOVA was justified. The control group consisted of 9 male and 11 female teachers. The total number of students in the control group was 280 including 111 male and 169 female students.

4. Results

Based on the objectives of the study, three research questions were formulated the results of which are presented in this section. As for the first research question, a One-way ANCOVA was conducted to determine a statistically significant difference between the control and the experimental groups' achievement test controlling for their pretest.

The experimental group included the classes in which FA had been continuously implemented and the control group included the classes in which teachers had not been using FA as an integral part of their teaching.

Preliminary checks were conducted to ensure that the assumptions underlying ANCOVA had been met. One assumption is measurement of the covariate. This assumption specified that the covariate should be measured before the treatment or experimental manipulation begins. This is not tested statistically, but instead forms part of the research design and procedures. This assumption was met because, as mentioned, pretest on academic achievement (the covariate) was done prior to the study. Another

assumption was reliability of the covariate. This assumption which concerns the reliability of the covariate is also part of the research design, and involves choosing the most reliable measuring tools available. This assumption was also taken care of since, as reported in the instrument section, the internal consistency reliability of the covariate was checked by calculating Cronbach alpha for the test of academic achievement. The final assumption was homogeneity of regression slopes and equality of variances which concerns the relationship between the covariate and the dependent variable for each of the groups. There should be no interaction between the covariate and the treatment or experimental manipulation. This is done through statistical procedures. In this case, Table 1 shows the results for the test of equality of variances.

Table 1.

Levene's Test of Equality of Error Variances

<i>F</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
128.986	1	649	.000

Dependent Variable: PostAchievement

As it is apparent from the output of Table 1, variances seem not to be homogeneous ($p < 0.05$). However, according to Leech, Barrett, & Morgan (2005) if the assumption of homogeneity of variance had not be met (found significant), it would not be a major problem providing the cell sizes are equal (i.e., the largest group size is not more than 1½ times greater than the smallest group size). This is the case for two reasons, first, the ANCOVA statistic is a robust statistic and second, because of the way SPSS calculates the ANCOVA. Therefore, we were able to continue the procedure. Table 2 indicates the descriptive statistics for the two groups.

Table 2.

Descriptive Statistics for Achievement

FA	Mean	Std. Deviation	N
Experimental	17.5499	1.40638	371
Control	15.9250	2.67701	280
Total	16.8510	2.20229	651

Dependent Variable: PostAchievement

After adjusting for pre-intervention scores, there was a significant difference between the experimental and control groups on post-intervention scores on student academic achievement, $F(1, 649) = 128.98, p = 0.01$. The significance level was set at .05. Furthermore, based on the figure for partial eta squared, the effect of FA on achievement was nearly strong. Table 3 shows ANCOVA results for the dependent variable, in that, student academic achievement.

Table 3.

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>Sig.</i>	Partial Eta Squared
Corrected Model	421.556 ^a	2	210.778	50.013	.000	.134
Intercept	4449.970	1	4449.970	1055.873	.000	.620
PreAhivement	.261	1	.261	.062	.803	.000
FA	417.485	1	417.485	99.059	.000	.26
Error	2730.991	648	4.214			
Total	188008.000	651				
Corrected Total	3152.547	650				

Dependent Variable: PostAchivement

Based on the second research question, it was meant to know which component of FA had a more powerful effect on student achievement. In order to explore this question, a number of steps were taken. First, for the twenty teachers in the experimental group, each component mean score was sorted out from the largest to the smallest value. In other words, it was made clear which teachers utilized each component of FA more than the others. Teachers' scores for the first five highest scores for each FA component were considered the basis for achievement scores of the same teachers. The mean scores for different components of FA were then calculated. Table 4 shows the descriptive statistics for FA components used by teachers. Based on the descriptive statistics reported in Table 4 below, the mean scores for learning targets, self-assessment, and peer-assessment were more than other components. On the other hand, monitoring score was the lowest among FA components.

Table 4.

Descriptive Statistics for FA Components

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
LT	86	30.1395	1.51977	.16388	29.8137	30.4654	26.00	33.00
MON.	94	29.0745	2.31516	.23879	28.6003	29.5487	21.00	32.00
FEED	100	29.2900	2.34540	.23454	28.8246	29.7554	21.00	34.00
SELF	85	30.0824	1.72662	.18728	29.7099	30.4548	26.00	34.00
PEER	91	30.0110	1.56698	.16426	29.6846	30.3373	26.00	33.00
Total	456	29.6974	1.99022	.09320	29.5142	29.8805	21.00	34.00

In order to know if there were any meaningful differences between the students' scores on academic achievement, achievement scores for the students were compared to students' achievement in other groups. A One-way ANOVA was used to know if there were any meaningful differences between the five groups. Before that, preliminary tests had been done to know how it would be possible to compare the means. To test the homogeneity of variances Levene's test was conducted and the following output (Table 5) emerged.

Table 5.

Test of Homogeneity of Variances for FA Components

Levene Statistic	df1	df2	Sig.
5.571	4	451	.000

As Table 5 indicates, the test of homogeneity of variances was not significant. Therefore, the ANOVA test was conducted with Robust Tests of Equality of Means. Based on these tests, F test was significant. The results appear in Table 6.

Table 6.

Robust Tests of Equality of Means

	Statistic ^a	df1	df2	Sig.
Welch	5.419	4	224.809	.000
Brown-Forsythe	6.172	4	410.035	.000

a. Asymptotically F distributed.

As shown in Table 6 above, there were significant differences between the five FA components in improving students' achievement. Using the Welch statistic, it was found that $F(4, 224.80) = 5.41, p < .05$. Since the a priori alpha level were set at .05, it was concluded that the adjusted F ratio was significant. Since the p value is smaller than .05, the null hypothesis was rejected and proceeding to comparing the group means was justified. Table 7 shows that the difference was significant.

Table 7.

ANOVA Results for FA Components

	Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>Sig.</i>
Between Groups	91.430	4	22.857	6.026	.000
Within Groups	1710.807	451	3.793		
Total	1802.237	455			

Based on the ANOVA Post Hoc tests, there was a significant difference between clarifying learning targets, monitoring, and feedback. There was also a statistically significant difference between monitoring and self-assessment, and also between monitoring and peer assessment. Finally, a significant difference was found between feedback and self-assessment. Therefore, it was concluded that there were significant differences among FA components to improve student achievement. Table 8 shows the statistical figures for FA components.

To investigate the third research question, a One-way ANCOVA was conducted to determine a statistically significant gender difference between the control and the experimental groups' achievement test controlling for their pretest. The experimental group included the classes in which FA had been continuously implemented and the control group included the classes in which teachers had not been using FA as an integral part of their teaching. Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate. Table 9 shows the results for the test of equality of variances.

Table 8.

Multiple Comparisons Based on Tukey Test for FA Components

Tukey HSD						
(I) group	(J) group	Mean Difference (I-			95% Confidence Interval	
		J)	Std. Error	Sig.	Lower Bound	Upper Bound
LT	MON.	1.06507*	.29063	.003	.2691	1.8610
	FEED	.84953*	.28643	.026	.0650	1.6340
	SELF	.05718	.29789	1.000	-.7587	.8730
	PEER	.12855	.29291	.992	-.6737	.9308
MON.	LT	-1.06507*	.29063	.003	-1.8610	-.2691
	FEED	-.21553	.27980	.939	-.9819	.5508
	SELF	-1.00788*	.29152	.005	-1.8063	-.2095
	PEER	-.93652*	.28643	.010	-1.7210	-.1520
FEED	LT	-.84953*	.28643	.026	-1.6340	-.0650
	MON.	.21553	.27980	.939	-.5508	.9819
	SELF	-.79235*	.28734	.048	-1.5793	-.0054
	PEER	-.72099	.28217	.081	-1.4938	.0518
SELF	LT	-.05718	.29789	1.000	-.8730	.7587
	MON.	1.00788*	.29152	.005	.2095	1.8063
	FEED	.79235*	.28734	.048	.0054	1.5793
	PEER	.07136	.29379	.999	-.7333	.8760
PEER	LT	-.12855	.29291	.992	-.9308	.6737
	MON.	.93652*	.28643	.010	.1520	1.7210
	FEED	.72099	.28217	.081	-.0518	1.4938
	SELF	-.07136	.29379	.999	-.8760	.7333

*. The mean difference is significant at the 0.05 level.

Table 9.

Levene's Test of Equality of Error Variances

<i>F</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
1.713	1	649	.191

Dependent Variable: PostAchievement

The significance level was set at .05. As it is apparent from Table 9, the assumption of homogeneity of variances is met because p (.191) is greater than α (0.05). Therefore, continuing with ANCOVA was statistically justified. It was concluded from the output that

there was not a significant difference between the experimental and control groups on post-intervention scores on student academic achievement with regard to students' gender, $F(1, 649) = 0.233$, $p = 0.630$. In other words, gender was not a determining factor in students' academic achievement following FA implementation by the teachers. Table 10 indicates detailed statistical figures.

Table 10.

Tests of Between-Subjects Effects

Dependent Variable: Post Achievement								
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^b
Corrected Model	5.201 ^a	2	2.600	.535	.586	.002	1.071	.139
Intercept	4921.579	1	4921.579	1013.293	.000	.610	1013.293	1.000
PreAchievement	3.927	1	3.927	.809	.369	.06	.809	.146
sGender	1.130	1	1.130	.233	.630	.000	.233	.077
Error	3147.346	648	4.857					
Total	188008.000	651						
Corrected Total	3152.547	650						

a. R Squared = .002 (Adjusted R Squared = -.001)

b. Computed using alpha = .05

5. Discussion

An analysis of data for the first research question revealed that ongoing implementation of FA had a significant impact on students' academic achievement. This implied that the teachers who use FA strategies effectively in their classrooms can expect their students perform better in their achievement tests. As for the second research question, findings indicated that, among the five components of FA, teachers in the experimental group used *clarifying learning targets* more than the others. The second most frequent FA strategy utilized by teachers was *self-assessment* and *Peer-assessment* stood at third place. This implied that most of the teachers preferred to reassure the students of what they were expected to do and make learning targets clear for them. They also tried to encourage students to have self- and peer-assessment as the criteria for success. *Giving feedback* and *monitoring* were the least FA strategies practiced by teachers in the experimental group. This implied that teachers found them less effective. The reason may be that the first three components are more student-centered.

The findings of this study are in line with past research findings on FA which were conducted by Wiliam, Harrison, and Black (2004) who reported a positive effect of FA on students' achievement. Similarly, Cizek (2010) confirmed the findings by reporting that FA is a valuable source from which both teachers and students can benefit. This study may have implications for both teachers and students in that, the teachers can use FA in their instructional planning and the students can deepen their understanding and improve their achievement. Additionally, Smith (2008) found that using FA effectively improved students' achievement that is in accordance with the findings of the present study. Finally, the findings of the present study confirmed what DeNome (2015) and Kline (2013) had found. They reported an achievement improvement in students whose teachers had received training in FA. The difference was the subject under investigation. Whereas this study investigated English language, in above-mentioned studies the effect of FA on mathematics and reading was investigated. Finally, the findings of the present study supported those of Kline (2013) indicating that gender does not play a role in the effect of FA on students' achievement.

6. Conclusion

This study contributes to literature supporting the positive instructional implications associated with FA. However, the study primarily contributes to literature that suggests high school students will be benefitted from FA practice in their classrooms by performing better in their class assessments. Because one of the main instructional goals for teachers and other people involved in the job of teaching is learner achievement, the present study indicated that this goal can be reached if FA is successfully implemented by teachers in classrooms.

Another implication which can be attributed to this study is the way FA was expected to be implemented. Throughout the study, FA was described as an ongoing and continuous process, which was an integral part of teachers' instructional practice. These attributes were used deliberately since FA can take effect if it is practiced as described. Otherwise, we cannot claim that we are using FA successfully. Therefore, the kind of FA we expect to lead improvement in students' achievement has to be an integral part of teaching practiced. This essential point was taken care of in this study by a number of observations in classrooms.

Furthermore, it became clear from the initial stages of data collection that a large

group of teachers do not use FA in their classrooms. Part of the problem may be attributed to teachers' lack of necessary knowledge or their misconception of FA. In other words, teachers may not be assessment-literate enough in conducting FA. They may also think they are practicing FA in their classes while they are not doing so in reality. This problem can be alleviated by conducting workshops and/or in-service courses on assessment in general and on FA in particular. The data collected in the classes taught by teachers in the experimental group confirmed this. Accordingly, the teachers were able to use FA strategies successfully and became confident enough to use FA efficiently after the workshop. In other words, professional development helped the teachers in the experimental group improve their FA literacy. This knowledge, in turn, helped them to implement FA successfully. It is suggested that in-service training courses for teachers be held to provide them with necessary knowledge needed to implement FA in their classes.

As Stiggins (2005) emphasized, improving FA literacy is important because some teachers want to use FA in their classrooms but do not know much about the subject. Therefore, it would be a good idea that the education organizations offer FA courses for teachers and welcome collaborative works on FA. Teacher's participation has to be encouraged, too. This depended on the education system, which can use variety of ways to motivate teachers to take part in such courses.

References

- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.
- Black P, & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7-74.
- Black, P, & Wiliam, D (1998b). *Inside the black box: Raising standards through classroom assessment*. London, UK: King's College London School of Education.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology, 78*(4), 210-216.

- Cizek, G., J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade and G. J. Cizek (Eds), *Handbook of formative assessment* (pp. 41-58). Oxford: Routledge.
- Davidheiser, S. A. (2013). *Identifying areas for high school teacher development: A Study of assessment literacy in the Central Bucks school district*. Unpublished doctoral dissertation, Drexel University, USA.
- DeNome, S., E. (2015). *The impact on student achievement following professional development on the principles of formative assessment*. Unpublished doctoral dissertation, William Howard Taft University. USA.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Research and Education*, 14(7), 1-11.
- Fisher, D., & Frey, N. (2007). *Checking for understanding: Formative assessments for your classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Calif.: Corwin.
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. L. (2009). From evidence to action: A seamless process of formative assessment? *Educational Measurement: Issues and Practice*, 28(3), 24-31.
- Kline, A., J. (2013). *Effects of formative assessment on middle school student achievement in mathematics and reading*. Unpublished doctoral dissertation, University of North Carolina. USA.
- Klute, M., Apthorp, H., Harlacher, J., Reale, M., & Research, M. (2017). *Formative assessment and elementary school student academic achievement: A review of the evidence*. Institute of Education Sciences. U.S. Department of Education.
- Macgregor, G., Spiers, A. & Taylor, C. (2011). Exploratory evaluation of audio email technology in formative assessment feedback. *Research in Learning Technology*, 19(1), 39-59.

- McDowell, L. Wakelin, D., Montgomery, C and King, S. (2011). Does assessment for learning make a difference? The development of a questionnaire to explore the student response. *Assessment and Evaluation in Higher Education*, 36(7), 749-765.
- McManus, S. M. (2008). *Attributes of effective formative assessment*. Retrieved March 3, 2016 from the World Wide Web:
http://www.ccsso.org/Documents/2008/Attributes_of_Effective_2008.pdf
- McMillan, J. H. (2010). The practical implications of educational aims and contexts for formative assessment. In H. L. Andrade and G. J. Cizek (Eds), *Handbook of formative assessment* (pp. 41-58). Oxford: Routledge.
- Oswalt S. G. (2013). *Identifying formative assessment in classroom instruction: Creating an instrument to observe use of formative assessment in practice*. Doctoral dissertation, Boise State University, USA.
- Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using the SPSS program* (4th ed.). Allen & Unwin, Berkshire.
- Popham, J. W. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Smith, L., W. (2008). *Using formative assessment results to predict student achievement on high stakes tests*. Unpublished doctoral dissertation, Liberty University, USA.
- Stiggins, R. J. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328.
- William, D. (2008). International comparisons and sensitivity to instruction. *Assessment in Education: Principles, Policy, and Practice*, 15(3), 253-257.