



Journal of Business and Social Science Review
Issue: Vol. 1; No. 2 February 2020 pp.73-84
ISSN 2690-0866(Print) 2690-0874 (Online)
Website: www.jbssrnet.com
E-mail: editor@jbssrnet.com

The Effect of Formative Assessment on the Academic Achievement Levels of Prospective Teachers

Davut HOTAMAN

Yıldız Technical University

Faculty of Education

Davutpasa Campus, 34220, Esenler, İstanbul, Turkey.

E-mail: davuthotaman@gmail.com

+90 212 3834827

Abstract

This study aims to examine the effect of “Formative Assessment (FA)” practices in “Assessment and Evaluation in Education” class on the academic achievement levels of prospective teachers. It uses a mixed research design. Quantitative data were collected by using double pretest-posttest design, which is one of the complete experimental design structures; while a semi-structured Interview Form was used to collect the qualitative data. The study group consists of a total of 220 prospective teachers who participated in a “*Teacher Training Course*” in Faculty of Education in Yıldız Technical University during the 2017-2018 academic year. The data collection instruments included a 40-item multiple-choice achievement test (AT) chosen from a question bank in accordance with the course objectives and a semi-structured interview form. For the achievement test, reliability was achieved by the test-retest method ($r=.95$), and validity was secured by the “*analytical*” method based on expert opinion. Following the nine-week FA practices using the complete experimental double pretest-posttest research design, it was found that these practices (*What do I recall?* and *What have I learned?*) resulted in a significant difference in favor of the experiment groups. The responses of prospective teachers to the semi-structured interview form developed to collect qualitative data for the study were categorized into common themes, which demonstrate that the quantitative data are confirmed by the qualitative data regarding the FA practices. This harmony between the quantitative and qualitative data showed that FA practices (independent variable) are strong enough to affect the achievements of prospective teachers (dependent variable).

Keywords: *Education, curriculum, assessment, formative assessment.*

Introduction

Education is a way of acculturation and self-actualization for individuals and a process of creating similar individuals for societies. Assuming the function of training the society at large and young generations as its future, educational institutions can only become effective through their curricula. Curricula are defined as systematic educational contexts aiming to train a certain group of students within a certain time period. Education is an open system (Sönmez, 2009, p.1); individual and social needs constantly change, and the curricula designed in theory do not always yield the expected results, which stems from the need to transfer new needs to the curricula and the adopted ideological philosophy (pragmatism). In this sense, an education system is arguably a dynamic structure. All curricula consist of “*learning goals/outcomes, content/scope, learning experiences/educational contexts and assessment*”. *Learning goals* refer to the desired qualities (knowledge, skills, habits, values, beliefs etc.) that are aimed to be acquired by an individual and can indeed be acquired through education (Ertürk, 1984, p.109); while *content* encompasses the elements that explain the knowledge, skills and habits signified by the goals (Demirel, 2015). *Learning experiences*, on the other hand, have to do with the practical part of the curriculum and represents educational activities which strategically combine the learning input such as materials, methods, techniques and technologies that help students acquire the learning outcomes with the variables determining the quality of the learning service (Sönmez, 2009, p.7). *Assessment* is the final stage in organizing educational activities. Assessment results underlie the decisions of whether curriculum changes yield the desired outcomes (Oosterhof, 2003).

Education clearly requires planned activities due to various factors such as the nature of time, critical stages for the learner, and the desire to exclude unwanted situations. This is why it is essential to know whether educational activities are as effective as they are planned. All kinds of assessment and evaluation activities carried out accordingly can guide the reconstruction of the process. Thus, evaluation allows education to restore itself (Ertürk, 1984, p.109; Stiggins, 2003, p.199; Oosterhof, 2003; Atılgan, 2017). Assessment is a requisite for evaluation. Assessment is the act of identifying the extent to which an entity, phenomenon or situation possesses a certain quality (Turgut, 1997, pp. 8-12; Özçelik, 2013, pp.10-12; Moss and Brookhard, 2009, p.5). Therefore, assessment is the process of observing any variable and converting the results into numbers or symbols (Tekin, 2016; Baykul, 2000). Evaluation, on the other hand, is the process of comparing the numbers and symbols identified through assessment to a certain criterion and arriving at a conclusion (Turgut, 1997; Başol, 2016, p.256; Tok, 2007; Tekin, 2016, p.7; Bahar et al., 2015, p.3).

Formative Assessment

Regardless of the type of evaluation, assessment practices focus on course objectives. This process examines all objectives expected to be learned as part of a subject and unit (Özçelik, 2013, p.10-12; Torrance and Pryor, 2002, p.8; Irons, 2008, p.6). In formative assessment (FA), a student's learning shortcomings, errors, learning pace and difficulties are identified to revise the parts that require restructuring and reorganization. In this process, instructors develop and use monitoring scales that examine all cognitive behaviors regarding the unit or subject (Ertürk, 1984, p.113). Formative assessment aims not to grade students but rather to test students' learning and to correct any shortcomings or incorrect learning, if any (Cauley and McMillan, 2010; Crooks, 2001). According to Tekin (2016, p.25) "*formative assessment (FA)*" should be regarded as a part of the teaching process. FA practices which do not involve any grading constitute an important assessment type that tests the *quality of a teaching service*. This type of assessment mainly serves to identify learning shortcomings and difficulties concerning the taught/learned information at the beginning and/or end of every course, subject and unit in the course of teaching or during the education process (Moss and Brookhard, 2009, p.5). Here, the aim is to instantly respond to any incorrect and incomplete practices without losing time in order to improve the quality of the teaching service (Özçelik, 2013, p.10-12; Crooks, 2001; Fisher, and Frey, 2007, p.120). As a result of such assessment, learning shortcomings and difficulties are identified, and initiatives are taken as a remedy (Karaca, 2010, p.24; Lin and Gronlund, 1995).

The teaching process is planned in order to identify learning shortcomings and difficulties. In determining objectives for each subject and unit, screening tests with a high content validity should be developed and administered at the beginning and/or end of every unit or chapter (Senemoğlu, 2007; Özçelik, 2013). This would help identify which learning outcomes are related to learning shortcomings and difficulties within a chapter or unit for each and every student (Frey and Fisher, 2011, p.119). Thus, the instructors will have the chance to see the effectiveness of their teaching service, and may remedy any shortcomings, incomplete learning and learning difficulties through additional teaching practices based on the screening results, which would in turn improve *the quality of the teaching service* (Bahar et al., 2015, p.4; Cauley and McMillan, 2010; Lin and Gronlund, 1995).

In addition to identifying learning and teaching shortcomings, FA practices also provide information about curricular effectiveness, satisfaction levels of students, positive or negative attitudes toward the curriculum and compatibility of the course with the theoretical framework (Başol, 2016, p.277; Irons, 2008, p.52). It is very important for the screening tests administered as part of the FA practices to represent the curricular content. Otherwise, such assessment practices would be superficial and haphazard, with no apparent benefits, but also result in a waste of time. Effective time management is vital in teaching as elsewhere. FA practices are informative and intended to improve, monitor, shape and develop the teaching-learning process. It is process-oriented, and an integral part of teaching activities.

Study Purpose and Significance

Assessment practices contribute to assessing the system, teacher and the student through the feedback they provide at every step of education process.

This general function of assessment is consistent with its mission but insufficient on its own. Assessment *may contribute to the learning process* as a variable defining the quality of teaching service. This study aims to identify whether academic achievement of prospective teachers are influenced by “*Formative Assessment (FA)*” practices, a type of assessment that intends not to grade students but rather develop and improve teaching service. The present study is also important in considering the function of assessment in terms of *improving the quality of teaching service*.

The study seeks answers to the following questions:

1. Does effective use of formative assessment practices (*tests aiming to identify the learning success, learning shortcomings and difficulties of students*) significantly impact upon the academic achievement of prospective teachers?
2. What are the opinions of prospective teachers about the effective use of formative assessment practices (*tests aiming to identify the learning success, learning shortcomings and difficulties of students*)?

METHODOLOGY

In line with the “compatibility thesis” which argues that “*quantitative and qualitative methods are compatible and can be used together in a single research*” (Balci, 2007, p.44), this study employs quantitative and qualitative methods together. Using both qualitative and quantitative methods, mixed methods ensure the insight and understanding which might be overlooked in case a single method is used (Balci, 2009, p. 44; Plano-Clark and Creswell, 2015; Creswell, Plano-Clark, Gutmann and Hanson, 2003; Giannakaki, 2005; Greene, 2005; Leech and Onwuegbuzie, 2009, p.267; Tashakkori and Teddlie, 2003), and in doing so, they enable to examine cases related to the dependent and independent variables, in depth (Best and Kahn, 2003, p.171; Creswell and Plano-Clark, 2017; Creswell, Clark, Gutmann and Hanson, 2003).

Of true experimental designs, *pretest-posttest design with double control groups* was employed in order to use quantitative data for the first question of the study (Köse, 2010, p.116; Karasar, 2012, p.98; Berg and Lune; 2015; Daley and Onwuegbuzie, 2004, p.552). Experimental research aim to test the effect of the differences generated by the researcher upon a dependent variable. The common characteristics of such designs include their use for multiple groups, group formation through unbiased assignment (sampling) and their suitability for multiple comparisons (Plano-Clark and Creswell, 2015; Karasar, 2012, p. 97). The study aimed to ensure double check for the “*experimental procedure*” by using double experiment and double control groups, and through a comparison using a wider sample. Table 1 presents the experimental design for the study.

Table 1: *Experimental Design for the Study*

Group	Pretest	Nine-week “WR” and “WL” Screening Tests									Posttest
Experiment-1 (E1)	Achievement test	X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	Achievement test
Control-1 (C1)	Achievement test	Traditional instruction									Achievement test
Experiment-2 (E2)	Achievement test	X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	₁ X ₂	Achievement test
Control-2 (C2)	Achievement test	Traditional instruction									Achievement test

₁=What do I recall (WR)?, ₂=What have I learned (WL)?

As seen in Table 1, the control groups received no treatment and the traditional course organization maintained while the two experiment groups were subjected to the experimental procedure. The “What do I recall(₁)?” screening test administered at the beginning of the course aims to monitor what students recall in their long-term memory regarding the previous lesson/subject and 8-10-item short screening tests are carried out according to the course objectives of previous lessons. The class starts with the “What do I recall?” screening test results and is built upon the daily subject. At the end of the class, the students received the “What have I learned (₂)?” screening tests which involve 8-10-item multiple choice questions that test the course objectives of the day. As it is stated above, the FA practices continued effectively for the period of nine weeks. The study is based on the mixed research design which is supported by qualitative and –primarily– quantitative data. From this point of view, it can be stated that explanatory research design is used in this study.

This sort of a method allows the researcher to focus on a certain case and gather detailed information by using descriptive questions (Johnson and Onwuegbuzie, 2004; Wellington, 2000). In the study, the qualitative data were collected by using a semi-structured four-item interview form and the collected data were analyzed through descriptive analysis. As the effect of experimental procedure on the dependent variable is presented quantitatively, students' views on the procedure is gathered qualitatively. Thus, quantitative and qualitative research methods are integrated with an explanatory approach (Heppner et al., 2013, p.292). Accordingly, it is tested how the systematic changes in the independent variable affect the dependent variable (Best and Kahn, 2003, p.171). No other variable is considered to influence the dependent variable. Thus, any changes in the dependent variable, whether realized or unrealized, stem from the independent variable.

Study Group

Selection of the study participant was based on *purposive sampling*, a type of sampling method that is not based on probability (Fraenkel and Wallen, 2008). The present study was carried out on four student groups who were enrolled in the "Assessment and Evaluation in Education" course among the 17th-term groups taking the "Pedagogical Training Course" offered by Faculty of Education in Yıldız Technical University during the spring semester of the 2017-2018 academic year. In selecting the study groups, the classes with high attendance rates were determined as the sample. The study group consists of students who come from different departments and faculties than Faculty of Education and female students make up the majority of the group. Two of these classes were randomly assigned as the experiment groups, while two of them were assigned as the control groups. Table 2 presents the results of the unpaired t-test performed on the pretest results concerning the equivalence of the experiment and control groups.

Table 2: Results of the t-test between the Pretest Scores of the Experiment and Control Groups.

Groups	Pretest (x)	N	t-test	p
Experiment G ₁ (E ₁)	24.18	55	.12	.91
Control G ₁ (C ₁)	25.50	55		
Experiment G ₂ (E ₂)	24.07	55	1.08	.28
Control G ₂ (C ₂)	25.25	55		
Experiment1pr+Experiment 2pr	24.12	110	1.82	.07
Control1pr+Control2pr	25.38	110		

P<.05

As shown by Table 2, no significant difference was found between the mean pretest scores for the experimentG1 and the control group C1 ($t= .12, p>.05$). Similarly, Table 2 also shows that there is no significant difference between the mean pretest scores of experimentG2 and controlG2 ($t= 1.08, p> .05$). Furthermore, no significant difference was found between the mean pretest scores of the two experiment groups and the two control groups ($t= 1.82, p> .05$). Equivalence was procured among the groups.

Data Analysis

It is the test of normality which identifies the particular statistical techniques to be used in the study. However, although a distribution is regarded as normal in study groups with data exceeding thirty (30) according to the central limit theorem (Newbold et al., 2017, pp.234-238), the researcher also performed skewness and kurtosis tests on the data pertaining to the pretest experiment and control groups, with the results given in Table 3 below.

Table 3: Pretest X and Sx Values, Skewness-Kurtosis Normality Values and Demographic Data for the Study Groups

Groups	Pretest		Skewness	Kurtosis	Gender		N	%
	X	S			Female	Male		
ExperimentG ₁ (e ₁)	24.18	4.99	.031	-1.197	19	36	55	50
ControlG ₁ (c ₁)	25.50	5.41	.565	.068	27	28	55	
ExperimentG ₂ (e ₂)	24.07	6.52	.727	-.118	31	24	55	50
ControlG ₂ (c ₂)	25.25	3.84	-.380	.965	35	20	55	
Total					112	108	220	100

According to the demographic data for the sample groups, it is seen that the number of male and female participants in the groups are more or less equal (Male $n=108$, Female $n=112$). The pretest arithmetic averages for the groups are also roughly equal ($X=e_{1pr}+e_{2pr}=24.12$ and $X=c_{1pr}+c_{2pr}=25.37$). As for the equivalence of the groups, skewness coefficients are .031 for Experiment G_{1pr} , .727 for Experiment G_{2pr} , .565 for Control G_{1pr} , and -.380 for Control G_{2pr} . The kurtosis coefficients were found to be -1.19 for Experiment G_{1pr} , -.118 for Experiment G_{2pr} , .068 for Control G_{1pr} , and .965 for Control G_{2pr} . Although any value between -1 and +1 as skewness and kurtosis values for any data set are considered to be acceptable for normal distribution, values between -2 and +2 are also usually acceptable (Arapkirlioglu and Tankız, 2011).

Data Collection Instruments and Their Development

This study employs a “multiple choice achievement test” and a “semi-structured interview form”.

Achievement Test (AT)

As a data collection instrument, the 40-item achievement test is used, which was taken from a question bank of the related lesson and has difficulty index values between $p=0.20$ and $p=0.80$, and eliminative difficulty index values between $r_{jx}=0.36$ and $r_{jx}=0.20$. When the 40-item achievement test was applied to 204 students in different groups taking the same class, the internal consistency reliability (KR-20) was found to be $r=0.89$. Because of the value for each correct question is 1, for each wrong and unanswered question is 0. As eliminative and difficulty index of items that constitute the test are different, the correction formula was not used. Instead, KR- 20 was preferred as reliability value. An “analytical” validity study was performed based on expert opinions for the validity of the scale (Balcı, 2009, p.110; Büyüköztürk, 2007, p.168). A researcher and two other academicians, all fields’ experts, assessed the number of questions and their distribution across subjects, and concluded that “the scale used has a good item/question sampling for the objective/learning outcomes to be assessed”. The two “Assessment and Evaluation Specialist” academicians assessed the questionnaire for each item using a two-choice item/question assessment form in which they responded as “suitable/valid” or “unsuitable/invalid” for each question. The compatibility percentage for this interview form was calculated to be .91 and this value was regarded as “the content validity based on expert opinions” (Balcı, 2009, p.110; Büyüköztürk, 2007, p.168).

Semi-Structured Interview Form (SSIF)

Below are the four (4) questions in the “Semi-Structured Interview Form (SSIF)” generated by consulting “specialist academicians” and formulated to identify the opinions of the prospective teachers in the experiment groups (110 individuals) concerning “Formative Assessment (FA)”, which constitutes the second problem of the study?

1. What do you think about *Formative Assessment Practices*?
2. To what extent has your academic achievement been affected by *Formative Assessment Practices*?
3. In what way did the *Formative Assessment Practices* influence the identification of your learning difficulties and shortcomings?
4. In what way did *Formative Assessment Practices* influence your study habits?

Procedure

This study examines the effects of “Formative Assessment (FA)” practices (independent variable) upon the academic achievement levels of prospective teachers (dependent variable). Thus, a total of four (4) groups were selected, including two experiment (e_1 and e_2), and two control groups (c_1 and c_2). The courses given to the experiment groups included “What do I recall (WR)?” screening tests at the beginning of each class and “What have I learned (WL)?” screening tests at the end of each class, all of which make up the “Formative Assessment (FA)” practices. The control groups, on the other hand, received traditional course instruction. Therefore, “assessment was used as a variable defining the quality of the teaching service” in the study. The courses given to the experiment and control groups were all given by the researcher as it is believed to prevent the effect of the *teacher factor* upon the dependent variable. The WR and WL screening tests which constitute the FA practices as part of the experimental treatment were prepared by the researcher in view of the weekly course content. Research treatment was continued two hours a week for a period of nine weeks.

RESULTS

The study sought answers to the following questions: “Does effective use of formative assessment (FA) practices significantly impact upon the academic achievement of prospective teachers?” and “What are the opinions of prospective teachers about the effective use of formative assessment practices (FA)?” The study was carried out with four student groups to seek answers to these questions. The researcher obtained the statistical results for the quantitative data pertaining to the posttest results of the *achievement test* administered to the two experiment and the two control groups, as well as the analysis results for the information obtained through a semi-structured interview form administered for qualitative data. A t-test was performed to determine whether there is any significant difference between the mean posttest scores of the experiment and control groups, and the results were given in Table 4.

Table 4: T-test Results for the Differences between the Mean Posttest Scores for the Experiment and Control Groups as a result of FA Practices.

Assessment (FA)	N	x	S _x	S _d	t	p
ExperimentG1po=e1po	55	79.56	5.87	54	5.51	.00
ControlG1po=k1po	55	69.49	8.79			
ExperimentG1po=epo1	55	79.56	5.87	54	10.05	.00
ControlG2po=c2po	55	69.63	5.28			
ExperimentG2po=epo2	55	77.69	6.72	54	5.51	.00
ControlG1po=c1po	55	69.49	8.79			
ExperimentG2po=epo2	55	77.69	6.72	54	6.75	.00
ControlG2po=c2po	55	69.63	5.25			
ExperimentG1+ExperimentG2=e1po+e2po	110	78.62	6.35	109	9.95	.00
ControlG1+ControlG2=c1po+c2po	110	69.56	7.22			

P<0.05, FA=Formative Assessment Practices.

As seen in Table 4, the *experimental procedure*, i.e. the initial FA practice of “What do I recall?” and the final FA practice of “What have I learned?”, resulted in differences in favor of the experiment groups in all comparisons between the experiment and control groups ($t_{(54)}=5.51$ between e1po and c1po, $p<0.01$; $t_{(54)}=10.05$ between epo1 and c2po, $p<0.01$; $t_{(54)}=5.51$ between epo2 and c1po, $p<0.01$; and $t_{(54)}=6.75$ between e2pr and c2po, $p<0.01$). The final comparison in Table 4 was made between the mean posttest scores of the experiment and control groups, which revealed that the experimental procedure resulted in a significant difference in favor of the experiment groups, as is the case in all other comparisons ($t_{(109)}=9.95$ between the mean e1po+e2po and c1po+c2po, $p<0.01$). Even though, acquired t-score came out as significant, cohen d which is described as the influence quantity, is also considered ($d=1.33$, common language effect size .95 with confidence interval .827). It can be interpreted that the obtained cohen d value shows that the difference originates from not coincidences, but from reality (the applied experimental practise) (Kalaycı, 2005).

Qualitative Data Analysis

The qualitative data in this study is gathered by analyzing the answers of students in the experimental procedure on the semi-structured interview form which presented their personal opinions about FA practices. Out of the student groups that the experimental procedure was applied, 103 students filled out the form. The answers of these 103 students on the semi-structured interview form which consisted of four (4) questions were analyzed one by one. Students’ answers on each question were examined with the “descriptive analysis” from fundamental analysis techniques. Their answers on the first (1st) question were classified within six (6) categories as (1)*Motivation*, (2)*Boredom*, (3)*Preparedness for Class*, (4)*In-Class Participation*, (5)*Recapitulation* and (6)*Other*” whereas the second (2nd), third (3rd), fourth (4th) and fifth (5th) questions were grouped within three (3) categories as “(1) Positive, (2) Negative and (3) Neutral.” The descriptive analysis results of each question were tabulated and presented below in the order of the questions.

Question 1: What do you think about *Formative Assessment Practices*?

Table 5: Opinions of ExperimentG₁ and ExperimentG₂ Students Responding to Question 1 about FA Practices with their Percentage and Frequency Values

Common Theme	Similar and Common Student Statements	f	%
Motivation	-I was positively influenced by the instantaneous feedback about our learning products through the WR and WL tests.	27 (+)	26.21
Boredom	-I am bored by the initial WR and the final WL short tests.	14 (-)	13.59
Preparedness for Class	-The initial WR and the final WL short tests encourages me to prepare for and listen attentively to the class.	15 (+)	14.56
In-Class Participation	-The initial WR and the final WL short tests encourage me to prepare for, listen attentively and participate in the class.	19 (+)	18.44
Recapitulation	At the beginning of the class, the WR test in particular ensures recapitulation of the previous week's subject for me.	13 (+)	12.63
Other Opinions	-It is the first time I witness such a practice, -It feels different, I will not comment on it. -I think it takes too much time.	7(+/-) 5(+/-) 3(+/-)	6.79 4.85 2.91
Total	103	103	100

WR= What do we recall?, WL= What have we learned?

Table 5 shows that the first question in the semi-structured interview form aiming to collect qualitative data was responded by 103 students. In an attempt to classify all these opinions under common themes, it was seen that they could be categorized in 4 themes without excluding any of the opinions. f=27 positive (+) student opinions have been categorized under the first common theme "motivation", f=14 negative (-) student opinions under the second theme "boredom", f=15 positive (+) student opinions under the third theme "preparedness for class", f=19 positive (+) student opinions under the "in-class participation" theme, and f=13 positive (+) student opinions under the "recapitulation" theme. f= 15 student opinions (total 14.56%) which do not fall under any of the common themes were categorized under another theme called "other opinions".

Question 2: Has your academic achievement been affected by *Formative Assessment Practices*?

Table 6: Opinions of ExperimentG₁ and ExperimentG₂ Students Responding to Question 2 about FA Practices with their Percentage and Frequency Values

Common Theme	Similar and Common Student Statements	f	%
Positive	-Yes, though nothing seemed to change initially, I noticed I was better on the exam day.	23	21.70
	-Yes, it made a lot of difference.	21	19.81
	-Thanks to the "What do I recall" part administered at the beginning, I was able to repeat the information I forgot.	34	32.07
Negative	-I do not think so, I am already a successful student.	13	12.27
	-I believe that the FA practice negatively affected my success since it shortens the class time.	07	06.60
Neutral Opinions	-I have not understood anything.	03	02.84
	-It might have increased or decreased my success.	05	04.71
Total	106	106	100

WR= What do we recall?, WL= What have we learned?

As shown in Table 6, 106 students responded to the semi-structured interview form generated to collect qualitative data. The results are classified in *three themes*, i.e. "positive", "negative" and "neutral". The number of students stating that the pre-class and post-class FA practices positively impacted on their academic achievement was f=78, in other words, 73.58% of the student population stating their opinions indicated that FA practices *positively* (+) affect their academic achievement. The number of students openly stating that their academic achievement was *negatively* (-) affected was f=7. However, including those students responding as "I am already a successful student" in the negative opinion group, it could be argued that f= 20 students, about 18.87% of all students, stating their opinions indicated *negative* (-) opinions. Students with neutral opinions are f=8, constituting 7.55% of all respondents.

Question 3: Did the *Formative Assessment Practices* influence the identification of your learning difficulties and shortcomings?

Table 7: Opinions of ExperimentG₁ and ExperimentG₂ Students Responding to Question 3 about FA Practices with their Percentage and Frequency Values

Common Theme	Similar and Common Student Statements	f	%
Positive	-Yes, at the end of the class I noticed that I had difficulty in learning certain subjects.	16	16.84
	-Yes, if it was not for the daily assessment types of “ <i>What do we recall?</i> ” and “ <i>What have we learned?</i> ”, then I would not have noticed that my learning of some subjects was incomplete or incorrect.	21 08	22.10 08.43
	-I believe they have contributed by all means.		
Negative	-I do not think that FA practices identified my incorrect learning.	13	13.68
Neutral Opinions	-Recapitulation was more important than our incomplete or incorrect learning.	22 15	23.16 15.79
	-They have significantly contributed to reminding us of the subjects we forgot.		
Total		95	95 100

As shown by *Table 7*, a total of 95 students responded to the Question 3 in the semi-structured interview form used to collect qualitative data. F= 45 students, around 47.37% of all the respondents, indicated *positive* (+) opinions, while f=13 students, 13.68% of the respondents, stated *negative* (-) opinions. f=37 students, 38.95% of the respondents, indicated neutral opinions for Question 3.

Question 4: Were *Formative Assessment Practices* effective in helping you revise your study habits?

Table 8: Opinions of ExperimentG₁ and ExperimentG₂ Students Responding to Question 4 about FA Practices with their Percentage and Frequency Values

Common Theme	Similar and Common Student Statements	f	%
Positive	-Yes, although we did not gain anything in the form of grades, they consolidated our habit of preparing for the class.	27	27.84
	-We now care more about our in-class feedback.	15	15.46
	-We have begun to spend more time to revision at home or work.	12	12.37
Negative	-In-class recapitulation made the recapitulation at home insignificant.	13	13.40
	-We thought that in-class practices were sufficient for success.	19	19.59
Neutral Opinions	-The practice did not have an effect on me at all, whether positive or negative, since I am a hard worker.	11	11.34
Total		97	97 100

As revealed by *Table 8*, a total of 97 students responded to the fourth and last question of the semi-structured interview form developed to collect qualitative data. f=54 students, 55.67% of all the respondents, indicated *positive* (+) opinions, stating that the FA practices positively influenced their study habits. On the other hand, f=32 students, 32.99% of all respondents, stated *negative* (-) opinions, indicating that the FA practices were not efficient enough to encourage them to revise their study habits. f=11 students, 11.34% of the respondents, were neutral, meaning that the FA practices did not have any *positive* or *negative* effect in changing their study habits.

CONCLUSION AND DISCUSSION

Despite the rapid global production of information in any field, the education processes and practices have not totally freed themselves from the traditional perspective (Özdemir, 2007, p.187; Eskicumalı, 2003, p.17). One of such traditional practices and perspectives concerns the fact that the function of assessment cannot go beyond assessing the *product* or *result*. If it is aimed to change educational fields in parallel with global developments, it is obvious that it should be started by revising the existing structure and reconsidering the keystones of the education system. Building upon this idea, the present study reconsiders the function assigned to assessment.

Such reconsideration of assessment should focus not on how student achievement is assessed, but on how assessment should be used to improve it (Stiggins, 2003, p.197; Cauley and McMillan, 2010; Crooks, 2001; Heritage, 2010, p.37; Fisher, and Frey, 2007, p.120). In addition to develop a new perspective toward assessment and opening it for discussion among educators, this study also aims to demonstrate the positive influence of using FA practices as a variable defining the quality of teaching service upon students' academic achievement scientifically.

Using two experiment and two control groups, this experimental study proves that effective FA practices are able to enhance students' academic achievements. The results of statistical analysis show that the FA practices caused a significant difference in favor of the experiment groups. This finding demonstrates that the FA practices used as a *teaching instrument* in this particular experimental procedure alone are able to influence the academic achievements of the experiment group students significantly. Thus, such assessment practices within the education process play a role that promotes learning (Bahar et al., 2015, p.4; Cauley and McMillan, 2010). In the light of the quantitative results of the study, the FA practices were found to cause significant increase in the students' academic achievements in both experiment groups, when compared to the control groups who received traditional course instruction (*see Table 4*). This has been confirmed by both a single comparison of the experiment and control groups and a comparison between the mean academic achievement scores of the two experiment and the two control groups as a whole. The increase in academic achievement levels in favor of the experiment groups indicates the need to take FA practices seriously and put them into use by teachers to improve the quality of teaching service (Moss and Brookhard, 2009, p.5; Crooks, 2001).

The qualitative data gathered through the semi-structured interview form constitute the second aspect of the research. Qualitative data are employed to gather detailed information about a phenomenon that cannot be gathered through quantitative data and to integrate them with quantitative data. Our qualitative data includes the student responses to the four semi-structured interview questions addressed to the experiment group students. In the analysis of these data, it is seen that student opinions about FA practices (*see Table 5*) are grouped under the categories of *motivation, boredom, preparedness for class, in-class participation, and recapitulation* and *other* sub-themes and except for the *boredom* sub-theme, the students indicated positive (+) opinions about FA practices in all the categories. Consequently, it is seen that a large majority of the responding students seem to have been satisfied with the nine-week FA practice. The second question of the interview form examines whether FA practices influence the academic achievements of students, and about 74% of the respondents to this question indicated positive opinions (*see Table 6*), stating that the FA practices had an impact upon their success. This is confirmed by the increase in the academic achievements of the students in the experiment group as a result of the FA practices. Baykul and Turgut (2000) argue that providing students with occasional feedback about their learning levels would encourage them to revise themselves. Through such activities, students are given the chance to assess their own performance and thus, they notice their shortcomings (Başol, 2016, p.277; Frey and Fisher, 2011; p.1).

Another semi-structured interview question used to collect qualitative data aimed to determine the effectiveness of FA practices in identifying the learning difficulties and shortcomings of students through their opinions. Of all the respondents, 47.37% indicated that the FA practices had a positive contribution, while 33% stated that the FA practices were not effective in identifying learning difficulties. The rest of the students expressed neutral opinions. In the analysis of the opinions, it is seen that the number of students with positive opinions exceed that of students with negative opinions. As argued in the literature, FA practices aim not to grade students but to test their learning so that any incomplete or incorrect learnings can be identified to make necessary corrections before the next subject. (Cauley and McMillan, 2010; Crooks, 2001). The final question of the interview form aimed to identify the impact of FA practices upon the study habits of students. Around 56% of the respondents indicated positive opinions, stating that the feedback they received based on the FA practices were effective enough to encourage them to revise their study habits (duration, study habit, caring more about in-class feedback etc). Karaca (2010, p.24) states that learning shortcomings and difficulties are identified by using the data gathered through FA practices and the needed remedial actions are taken. The results of the present study show that FA practices influence student achievement positively. Therefore, "*formative assessment (FA)*" should be regarded as an integral part of the teaching process which has a favorable influence on learning level (Tekin, 2016, p.25; Başol, 2016, p.277; Frey and Fisher, 2011; p.1). Nevertheless, it is obvious that many teacher still seem to put an effort to improve the success of their students.

Still, they are not aware of the fact that the fundamental learning principle “*learning is based upon a previous one and serves as a preparation for the next*” (Yıldızlar, 2011, pp. 214; Sönmez, 2009) can be put into practice by using formative assessment during in-class teaching activities. In this sense, FA practices may contribute to the significant unity between learnings by identifying learning difficulties, recapitulating and refreshing previous learning through initial “What do I recall?” and final “What have I learned?” short screening tests about the previous lesson. The results of the present study show that there are ways to improve in-school learning and academic achievements of students, and teachers can improve desired learnings, and academic achievements accordingly, if they properly revise and activate certain implementations that they may have overlooked or neglected. The proper and continuous use of FA practices in the form of the initial (What do I recall?) and the final (What have I learned?) implementations would attribute a new perspective and mission to assessment, transforming it into an important teaching variable that defines the quality of teaching service in the form of “*hints, reinforcers, feedback-correction and student participation*”.

Suggestions

1. When the study results are considered, at the beginning and end of every lesson, a screening instrument should certainly be developed to perform short-time assessments with the function of “*reminding*” and “*refreshing*”.
2. In-class FA practices can bridge the gap in cases where the student fail to sufficiently repeat the lessons at home and spend time on studying.
3. By considering students’ level of tiredness, always focusing on grading in assessment may turn assessment, classes and school into a less pleasant setting.
4. FA practices aiming to identify the incorrect, incomplete learnings and learning difficulties of students instead of grading them may overcome the problems stated in item 3 above.
5. Using FA practices, the students have declared that they able to correct their own incorrect learnings, thus teachers should provide timely responses to reconstruct the assessment process without waiting for the end of the semester.
6. Through FA practices, students may have the chance to revise their own studying methods immediately.
7. Since FA practices give teachers the chance to make decisions about their own practices before waiting for the end of the semester, teachers should also revise their in-class practices.

References

- Ainsworth, L. A. & Viegut, D. (2006). *Common formative assessments*. Thousand Oaks, CA: Corwin Press.
- Arapkırli, H. & Tankız, K. D. (2011). Müzik öğretmenliği programı özel yetenek sınavlarında alan ve yerleştirme puanlarının karşılaştırılması (İnönü Üniversitesi Örneği). *e-international journal of educational research*, 2(4), 55-69, Autumn.
- Atılğan, H. (2017). Değerlendirme ve not verme, *Editör: Hakan Atılğan, Eğitimde ölçme ve değerlendirme (10. Baskı)*. Ankara: Anı Yayıncılık
- Bahar, M., Nartgün, Z., Durmuş, Ş. & Bıçak, B. (2015). *Geleneksel-tamamlayıcı ölçme ve değerlendirme teknikleri*. Ankara: Pegem Akademi.
- Başol, G. (2016). *Eğitimde ölçme ve değerlendirme (4. Baskı)*. Ankara: Pegem Akademi.
- Baykul, Y. (2000) *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM yayınları
- Best, J. W. & Kahn, J. V. (2003). *Research in education* (9. Edition). Boston: Allyn and Bacon.
- Büyüköztürk, Ş. Kılıç-Çakmak, E., Akgün, Ö.E., Karadeniz, Ş. & Demirel, F. (2013). *Bilimsel araştırma yöntemleri* (14. Baskı). Ankara: Pegem Akademi Yayıncılık.
- Cauley, K. M. & McMillan, J. H. (2010). *Formative assessment techniques*. The Clearing House.
- Chur-Hansen, A. & Koopowitz, L.F. (2005). Formative feedback in Teaching Undergraduate Psychiatry, *Academic Psychiatry*, 29(1):.66-69, Spring (<http://ap.psychiatryonline.org>)
- Crooks, T. (2001). The validity of formative assessments. *British Educational Research Association Annual Conference, University of Leeds, September*.
- Creswell, J. W. & Plano Clark. V. L.(2017). *Mixed methods research* (Third Edition). SAGE Publications Inc. US.
- Creswell, J.W. (2002). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson Education, Upper Saddle River, NJ.

- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. & Plano Clark, V. L., Gutmann, M., & Hanson, W. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (p. 209-240). Thousand Oaks, CA: Sage.
- Daley, C. E., & Onwuegbuzie, A. J. (2004). "Attributions toward violence of male juvenile delinquents: A concurrent mixed methods analysis". *Journal of Social Psychology*, 144, 549– 570.
- Ertürk, S. (1984). *Eğitimde program geliştirme (4. Baskı)*, Ankara: Meteksan.
- Eskicumalı, A. (2003). Eğitim ve toplumsal değişme: Türkiye'nin değişim sürecinde eğitimin rolü, 1923-1946. *Bogaziçi University Education Journal*, 19(2):15-29.
- Fisher, D. & Frey, N. (2007). *Checking for understand formative assessment Techniques for your classroom*. ASCD Alexandria, Virginia, USA
- Frey, N. & Fisher, D. (2011). *The formative assessment action plan*. ASCD Alexandria, Virginia, USA
- Giannakaki, M. S. (2005). "Using Mixed-Methods to Examine Teachers' Attitudes to Educational Change: The case of the Skills for Life Strategy for Improving Adult Literacy and Numeracy Skills in England". *Educational Research and Evaluation*, 11(4): 323- 348.
- Glesne, C. (2013). *Nitel Araştırmaya Giriş*. Ankara: Anı Yayıncılık.
- Greene, J. C. (2005). "The generative potential of mixed methods inquiry". *International Journal of Research & Method in Education*, 28(2): 207 – 211.
- Heppner, P. P., Wampold B.E, & Kivlighan D.M.(2008). *Psikolojik Danışmada Araştırma Yöntemleri*. Ankara: Mentis Yayıncılık.
- Heritage, M. (2010). *Formative assesment*. Cormin A SAGE Company, USA
- Irons, A. (2008). *Enhancing learning through formative assessment and feedback*. Taylor&Francis Group, New York
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). "Mixed methods research: A research paradigm whose time has come". *Educational Researcher*, 33(7): 14-26.
- Kalaycı, Ş. (2005). *SPSS uygulamalı çok değişkenli istatistikteknikleri*. (Editör: Ş. Kalaycı) Ankara: AsilYayımları.
- Karaca, E. (2010). Ölçme ve değerlendirmede temel kavramlar, Editör, Müfit Gömleksiz, Serdar Erkan, *Eğitimde Ölçme ve Değerlendirme (2. Baskı)*, Ankara: Nobel Yayın.
- Karasar, N. (2012). *Bilimsel araştırma yöntemleri (14. Baskı)*. Ankara: Nobel Yayın.
- Köse, E. (2010). Bilimsel araştırma modelleri (Edit.: Remzi Y. Kınçal), İçinde. *Bilimsel Araştırma Yöntemleri*, Ankara: Nobel Yayınevi.
- Leech, N. L. & Onwuegbuzie, A. J. (2009). "A typology of mixed methods research designs". *Qual Quant*. 43, 265–275.
- Lin, R. L. & Gronlund, N. E. (1995). *Measurement and assesment in teaching*. Upper Saddle River: Printice-Hall Inc.
- Moss, C. M. & Brookhard, S. M. (2009). *Advancing formative assessment in every classroom*. ASCD Alexandria, Virginia
- Newbold,P., Carlson, W. L. & Thorne, B. M. (2013). *Statistics for Business and Economics* (Eighth Edition). Pearson Education Inc. USA.
- Oosterhof, A. (2003). *Developing and using classroom assessments (Third Edition)*. Merrill Prentice Hall, Ohio, USA
- Senemoğlu, N. (2007). *Gelişim, öğrenme ve öğretim (Kuramdan Uygulamaya)*. Ankara: Gönül Yayıncılık
- Sönmez, V. (2009). *Öğretim ilke ve yöntemleri (3. Baskı)*. Ankara: Anı Yayıncılık
- Stiggins, R. J. (2003). *Assessment, student confidence and school success (Edit.: Allan C. Ornsteain, Contemporary Issues in Curriculum)*. Pearson Education, Inc.
- Özçelik, D. A. (2013). *Okullarda ölçme ve değerlendirme (öğretmen el kitabı) (2. Baskı)*, Ankara: Pegem Akademi.
- Özdemir, M.Ç. (2007). Toplumsal değişme karşısında aile ve okul. *Türk eğitim bilimleri dergisi*, Bahar, 5(2):185-198.
- Tashakkori, A. & Teddlie, C. (Eds). (2003). *Handbook of mixed methods in social and behavior sciences*. Thousand Oaks, CA: Sage.
- Tekin, H. (2016). *Eğitimde ölçme ve değerlendirme (26. Baskı)*. Ankara: Yargı Yayınevi.
- Tok, Ş. (2007). *Öğretme-öğrenme strateji ve modelleri*. Editör; Ahmet Doğanay, İçinde, *Öğretim İlke ve Yöntemleri*, Ankara: PegemA

Torrance, H. & Pryor, J. (2002). *Investigating formative assessment* (teaching, learning and assessment in the classroom). Open University Press, Philadelphia, USA

Turgut, M. F. (1997). *Eđitimde ölçme ve deęerlendirme*. Ankara: Nüve

Turgut, M. F. & Baykul, Y. (2012). *Eđitimde ölçme ve deęerlendirme (4. Baskı)*. Ankara: Pegem Akademi

Yıldırım, A. & Şimşek, H. (2011). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.

Yıldızlar, M. (2011). *Öğretim ilke ve yöntemleri (2. Baskı)*. Ankara: Pegem Akademi.