

Evaluating the Testing Effect in the Classroom: An Effective Way to Retrieve Learned Information

Elif ATABEK YİĞİT*

Fatime BALKAN KIYICI**

Gamze ÇETİNKAYA***

Suggested Citation:

Atabek Yiğit, E., Balkan Kıyıcı, F. & Çetinkaya, G. (2014). Evaluating the testing effect in the classroom: An effective way to retrieve learned information, *Eurasian Journal of Educational Research*, 54, 99-116.

Abstract

Problem statement: Evaluation, an important step in educational settings, is usually understood as a process to measure what students know or what they have learned. A variety of methods can be used for assessment and tests are one of the most important and widely-used. While being tested, one may learn or retrieve previously learned information via some mental processes that work on the memory. This phenomenon is called the “testing effect.” Despite some disadvantages, tests can also be used as learning materials. So, we will present our study on the testing effect in the classroom setting.

Purpose of study: The purpose of this study was to investigate whether the testing effect occurs in a classroom setting while using a test consisting of multiple choice and matching questions and a worksheet that summarizes the topic, and also to examine the effects of feedback and time.

Methods: In this study, the testing effect was investigated in a college chemistry course, and 98 pre-service science teachers participated. A pre-test, post-test, control group research design was followed to investigate the testing effect. A pre-test that has 100 short-answer questions was performed and students were grouped according to scores from that test. Seven groups (six experimental and one control) were constituted with the requirement that each group had the same average score on the pre-test. An intervening test was applied to four groups (two of them received feedback immediately after the test), a worksheet that summarizes the topic was studied by two groups and one group (control group) had no additional activity. The same pre-test was applied as a post-test to determine final retention. Three groups received this post-test a day later,

* Dr., Sakarya University College of Education, Sakarya, Turkey, eatabek@sakarya.edu.tr

** Corresponding Author: Dr., Sakarya University, Sakarya, Turkey, fbalkan@sakarya.edu.tr

*** Res. Assist. Middle East Technical University, Ankara, Turkey, cegamze@metu.edu.tr

and the other three experimental groups and the control group received it a week later. Final retention of previously learned information and the effects of testing, receiving feedback and re-studying were investigated.

Finding and Results: The results of this study showed that exposing students to supporting practices has a positive effect on retention of previously learned information regardless of the type of the practice. Specifically, tests, which educational professionals frequently use to assess their students' learning, should be used to support teaching and learning processes instead of just to determine the level of learning.

Conclusions and Recommendation: The results have important implications for classroom practice. That is, since much research supports the claim that testing has an important effect on students' retention of previously learned information, it, therefore, should be used to improve classroom practices, and support teaching and learning processes.

Keywords: Testing effect, feedback, retrieval, retention, science education

Introduction

Evaluation, an important step in educational settings, is usually understood as a process to measure what students know or what they have learned (Roediger & Karpicke, 2006a, 2006b; McDaniel, Anderson, Derbish, & Morrisette, 2007). A variety of methods can be used for assessment and tests are one of the most important and widely used. They are generally used because they require a shorter time for assessment (McDaniel, Roediger & McDermott, 2007). Although preparation of tests requires spending a lot of time, since they can be administered to large groups easily and scored objectively, they are frequently used (Chang, Yeh & Barufaldi, 2010). Besides, students may also prefer tests for evaluation. They may think that they have a chance to find the correct answer even if they do not have enough subject knowledge, which is true. One can choose the right answer in a test only by chance. This reality, actually, is a restriction, or a disadvantage of these evaluative materials. Moreover, tests restrict the ideas of students by giving them choices. Since students are forced to choose an answer from choices provided to them, they cannot express their own ideas or explanations about that topic (Mintzes, Wandersee & Novak, 2001).

Despite all of these disadvantages, tests can also be used as learning materials. While taking a test, one may learn or retrieve previously learned information via some mental processes that work on the memory. This phenomenon is called the "testing effect." Tests can enhance retention of previously learned information even if no additional study or feedback was provided, an effect investigated in many research studies, especially in the field of cognitive psychology (Roediger & Karpicke, 2006a).

In this study, we will first briefly summarize the literature about the testing effect, try to explain the mechanism behind this phenomenon and different variables

(different kinds of tests, delay between tests, feedback etc) to understand the effect under different conditions (psychology labs and classroom). Then we will present our study on the testing effect in the classroom setting.

In most of the studies, word lists (Carpenter & DeLosh, 2006; Cull, 2000; Wheeler, Ewers & Buonanno, 2003), animations (Johnson & Mayer, 2009), figure lists (Wartenweiler, 2011) or prose passages (Thomas & McDaniel, 2007; Agarwal, Karpicke, Kang, Roediger & McDermott, 2007) have been used as materials, and with a post-test their effect on retrieval was examined. Roediger and Karpicke (2006a) have investigated the testing effect through two experiments using prose passages. They created a study aiming to see the testing effect with one testing group versus one re-studying group. Their study also tried to determine the effect of time. It was concluded that re-studying enhances performance on immediate retention tests; however, testing has a more positive effect on delayed retention tests. They also concluded that repeated studying had a positive effect on an immediate retention test (5 min.), whereas repeated testing enhanced performance better on a delayed retention test. Wheeler, Ewers & Buonanno (2003), have also investigated the testing effect by comparing test trials and re-studying conditions. Their results shared the same pattern with many other investigations (Roediger, & Karpicke, 2006b; Butler & Roediger, 2007), that re-studying enhances retention in short intervals while testing enhances retention on delayed tests.

How the Testing Effect Occurs

One of the explanations of how the testing effect occurs is that additional exposure to learning material (the amount of processing hypothesis) enhances retention. However, many researchers have refuted this explanation in different studies (Carpenter & DeLosh, 2006; Roediger & Karpicke, 2006b) in which control groups were exposed to material (for instance, by re-studying) for the same amount of time as other groups spent being tested. Today, two main views are thought of as explanations of testing effect: the transfer-appropriate processing view and the elaborative retrieval processing view. According to transfer-appropriate processing view, the testing effect occurs because of the similarities between intervening and final tests. This explanation has found support in many research studies (Thomas & McDaniel, 2007; Butler & Roediger, 2007). A study by Wartenweiler (2011) showed that the testing effect can be explained by the transfer-appropriate processing view. He used figure lists as material and formed study-only and study-test groups. The testing effect, however, was only found to be significant for the transfer final test, not for the standard final test. In a study by Thomas and McDaniel (2007), prose passages were used as materials. Researchers gave two different types of passages to students, either letters were missed or sentences were disordered. Therefore, they wanted students to perform two types of encoding: letter insertion and sentence sorting. At the end of the study, it was argued that the testing effect occurred due to transfer-appropriate processing, since letter insertion encoding yielded better performance on the final cued recall test. Similarly, Johnson and Mayer (2009) have argued that the testing effect occurs according to the transfer-appropriate processing view.

The other explanation of the testing effect is that tests evoke more elaborative retrieval processing than studying. In other words, information that requires more mental processing leads to better retention when it is being tested. Several studies have supported this view (Carpenter & DeLosh, 2006; Wheeler, Ewers & Buonanno, 2003; McDaniel, Roediger & McDermott, 2007; Carpenter, 2009; Karpicke & Zaromb, 2010) and concluded that when an intervening test presents the information in a more complex way, participants' retention of that information on the final test will be improved. For this reason, free recall tests perform better than cued recall tests and they also perform better than recognition tests. Carpenter and DeLosh replicated Glover's fourth experiment in his study (1989; as cited in Carpenter and DeLosh, 2006) about the testing effect, and they investigated the elaborative retrieval processing view of the testing effect. Wheeler, Ewers and Buonanno (2003) examined the mechanism of the testing effect by using word lists as materials in two experiments; repeated study (multiple study trials without a test) and repeated test (a study trial followed by multiple recall tests) conditions and it was concluded that an item's storage strength would be increased by retrieval, and, therefore, it can be remembered easily. Instead of word lists, brief articles, lectures and materials were used as lecture materials in a college course used in a study by McDaniel, Roediger and McDermott (2007), and they found that an initial short-answer test produced greater gains on a final test than did an initial multiple-choice test.

Although there are two main explanations of the testing effect, it also should be noted that these explanations are not separated from each other with sharp lines. Both play a role in the testing effect.

Effect of Feedback

In some experiments the effects of feedback were also investigated. For instance, Agarwal, Karpicke, Kang, Roediger and McDermott (2007) examined the testing effect in open-book and closed-book tests, with and without feedback. The conclusion of their study was that providing feedback resulted in better long-term retention than providing immediate feedback. While in many investigations feedback was found to have a positive effect on final retention (Kang, McDermott & Roediger, 2007; Cull, 2000), a surprising result that feedback is ineffective has been found in a study by Butler and Roediger (2007). Video lectures were used in their study of three groups: a studying group (viewing lecture notes after watching video lecture), a short answer testing group and a multiple choice test group. Half of both testing groups were given feedback after testing while the other half were not. Retention of information was tested in a short-answer final test one month later. The surprising result in this study is that feedback had no effect on the final retention test. The researchers explained this result as due to the amount of time participants were given to process the feedback and the fact that it occurred immediately after subjects responded. Feedback was presented for only 6 seconds and this amount of time may not have been sufficient to allow participants to fully process the information.

Testing Effect Studies in Classroom

Most of the studies conducted on the testing effect have been done in psychology laboratories. However, in order to answer the question of whether tests are helpful to learning outcomes in a real classroom environment, it is required to study a classroom environment. Actually, there are many differences between the laboratory and the classroom. First of all, the amount of information that is to be learned by students is much more in the classroom than in laboratory designs. In the classroom, students may also differ in their attitude toward a lecture and in their motivation to learn the information. Every student requires a different amount of time to understand and learn material. Also, the materials to be learned are served in a variety of ways, such as textbooks, lectures, and classroom discussions. These differences between the classroom and the laboratory, and also the uncontrollable parameters in a classroom, make classroom studies harder to conduct than laboratory studies. (Roediger and Karpicke, 2006b)

However, the testing effect has been studied by some researchers in classroom settings. Bangert-Drowns et al. (1991) have studied whether tests affect learning outcomes in the classroom. With this aim, in many studies they grouped students in testing and no-test (control group) groups. Students in the testing group were administered a test during the semester, while the control group students did not have any test, but re-read lecture notes. Researchers investigated the testing effect by examining the students' final exam scores. They have found a positive effect of testing on the final score in 29 studies (of 35 studies), a negative effect in five studies, and no difference in one study. They concluded that the testing effect is also robust in the classroom. Bangert-Drowns et al. also investigated the number of tests in the testing group on the final performance. The number of tests taken in the testing group was changed between 3 and 75 while the control group received 0-15 tests. The results showed that as the number of tests increases, the positive effect on performance also increases. The important finding of the study is that the biggest difference in the effect of testing has been found when the control group had no test and the testing group had only one test. Therefore, they concluded that having only one test can produce better retention than a no-test condition. Although they found that tests are important tools and have a positive effect on final retention, they did not study the different kinds of tests or feedback conditions.

McDaniel et al. (2007) studied the testing effect in a web-based lecture course throughout a semester. As in many studies, they grouped students into a testing group and a re-studying group. McDaniel et al. used two different types of tests (multiple-choice and short-answer tests), which differed from Bangert-Drowns et al.'s study. The final exam scores of all students were examined, and they showed that students in the testing groups performed better than students in the re-studying group. From this result, researchers concluded that tests have positive retention effects. One other result from this study was that short-answer tests produce more gain than multiple-choice tests. This result has been paralleled in laboratory studies, in which recall tests produced more retention than recognition tests.

Another study revealing the testing effect in the classroom was done by Leeming (2002). He used an exam-a-day procedure, in which students take a test before every lecture, instead of four exams throughout a semester. He used this procedure for two lecture courses (Introductory Psychology and Learning and Memory) with 22-24 exams per course, and at the end of the semester students' final exam scores were compared to those in previous years. He concluded that students performed better when the exam-a-day procedure was used. Also, students' responses to a questionnaire related to the application of the procedure was analyzed, and it was discovered that students have positive attitudes toward this procedure and said they spent more study time and thought that they had learned more.

Another study dealing with the testing effect in the classroom was done by Chang, Yeh and Barufaldi (2010). Different from other studies, the participants of this study were primary school students (N=208), and the amount of retrieval was determined via the flow-map technique, a baseline instrument used to probe students' cognitive structures. Testing groups and a control group were constructed according to the scores that were obtained using a flow-map technique. A multiple-choice test, a correct-concept test and an incorrect-concept test were used as materials. Chang et al. concluded from the results of the study that tests led to better retention of learned material, and that, from a conceptual-change point of view, the increase in students' correct concepts stem from correct statements in a test, while incorrect statements may cause misunderstanding.

In the last decade, as part of educational reforms, new science education programs were prepared using a student-centered approach. These programs, which put students at the center of the system, have an evaluation method that supports learning activities and also gives feedback. It has great importance that process should be evaluated with outcomes as well according to educational reforms. However, if it was taken into consideration that in our country individuals were exposed to tests frequently and teachers use tests in their classes to evaluate students' performances, it is important that these evaluation materials should also be used in retrieving the learned information. Therefore, studies about the retrieval effects of testing materials are promising.

Another purpose of our study was to investigate the testing effect in a classroom setting using a test consisting of multiple-choice and matching questions and a worksheet that summarizes the topic (for re-studying). A pre-test that has 100 short-answer questions was taken and students were grouped according to scores from that test. Seven groups (six experimental and one control) were constituted, with the requirement that each group had the same average score on the pre-test. An intervening test was given to four groups (two of them received feedback immediately after the test), a worksheet that summarized the topic was studied by two groups and one group (control group) had no activity. The same pre-test was given as a post-test to determine final retention. Three groups received this post-test a day later and the other three experimental groups and the control group received it a week later. Therefore, final retention of previously learned information and the effects of testing, receiving feedback and re-studying were investigated.

Method

Research Design

This study was carried out as quasi-experimental design. Quasi-experiment includes assignment, but not random assignment, of participants to groups (Creswell, 2005). In this study, there were seven groups. One of them was a control group while others were experimental groups. The groups were equal on the pre-test score.

Participants

The participants of this study were 98 freshmen from the Elementary Science Education department of Sakarya University in Turkey. Of the participants, 30 were male and 68 were female, and all were enrolled in the General Chemistry I course. Before we conducted the study, they were all informed about the procedure and all of them participated voluntarily.

Research Instrument and Procedure

Pre-test and Post-test

A form consisting of 100 short-answer questions (which is enough to understand whether students learned the subject and to minimize their finding the right answers only by chance) about the naming of compounds (whether the name of the compound was given and the formula of it was asked or vice versa) was prepared and used as pre-test and post-test in this study. In 50 of the questions, the formula of the compound was given and the name of the compound was asked (e. g. Formula: Na_2SO_4 , Name=?), and in the other 50 the name was given and the formula was asked (e.g. Name: Potassium Chloride, Formula=?).

Intervening Test

An intervening test on naming compounds and consisting of two parts was used. In the first part, 10 multiple-choice questions on the rules of naming of chemical compounds were asked. The second part was composed of 100 matching questions in which students were asked to match the name and formula of a compound. Since it would be very confusing and difficult for students to find the right answer among 100 alternatives, this part was divided into 10 subparts composed of 10 questions in the same format. The names and formulas of 10 different chemical compounds were given without any order in the same section in two columns and participants were asked to match the name and the formula of a compound. All tests were examined by an outside chemistry specialist before administration.

Worksheet

A two-page worksheet, summarizing the topic with the basic rules of naming chemical compounds and examples, was used for re-studying practice. All basic rules of naming chemical compounds were summarized in this worksheet and examples of each rule were provided below the explanation. Students in re-studying groups studied this worksheet during the same time period as the test administration.

Design and Procedure

In this study, the naming of chemical compounds was chosen for the General Chemistry I course because this topic has great importance throughout this course and other chemistry courses as well. Students usually have difficulties in applying naming rules and learning this topic. It often seems like learning a foreign language. The study was conducted after that topic was taught in the classroom. All of the

participants attended the same lessons and were exposed to the same information on the topic by the same instructor.

A pre-test, post-test, control group research design was followed to investigate the testing effect of different interventions in a classroom setting. First, the pre-test was administered to all of the participants at the same time, and seven sub-groups (six experimental groups and a control group) with equal mean pre-test scores (33.29) were formed by entering the pre-test scores into an Excel spreadsheet. Six experimental groups (G1 to 6) were again divided into three different practice groups: Of the six experimental groups, two groups were administered the intervening test (G1 and G3), two groups took the intervening test and then received feedback (G2 and G4) and the other two groups (G5 and G6) studied the worksheets. The reason three different practice groups were formed was to be able to investigate whether there is a difference between the effectiveness of testing, receiving feedback and re-studying worksheets on retention of previously learned information, or if the testing effect occurs regardless of the type of study material. Since one of the earlier explanations of the testing effect is re-exposure to the material, a re-studying group was formed to examine whether this explanation is true or not. That is also why the control group was formed; to be able to see the effectiveness of all the practices.

Three weeks after the pre-test administration, the intervening practices were administered to the six experimental groups at the same time; four of the groups (G1 to 4) were administered the intervening test and two groups (G5 and G6) studied the worksheets for a class hour (50 minutes). The control group did not receive any interventions. At the end of that class hour, two of the groups (G2 and G4), which were exposed to intervening test administration, were given feedback immediately after the test; the instructor explained the correct answers to all of the questions in the test and supported these explanations with appropriate examples.

One day later, the post-test was administered to three practice sub-groups; a testing sub-group (G1), a testing with feedback sub-group (G2) and a re-studying sub-group (G5). The other three experimental sub-groups (G3, G4 and G6) and the control group (G7), which had no intervening activity, took the post-test one week later. In this way, the effect of time on retention of previously learned information was also to be investigated. A clear summary of the intervention program and time schedule of the study can be seen on Table 1.

Table 1

Time Schedule for Post-Test Administration

	Experimental groups						Control group
	1	2	3	4	5	6	
Intervening application	Test	Test + Feedback	Test	Test + Feedback	Worksheet	Worksheet	-
Post-test administration	1 day later	1 day later	1 week later	1 week later	1 day later	1 week later	1 week later

Scores on the post-test from all the experimental sub-groups and the control group were calculated and the results were analyzed. A schematic view of the study can be seen in Figure 1.

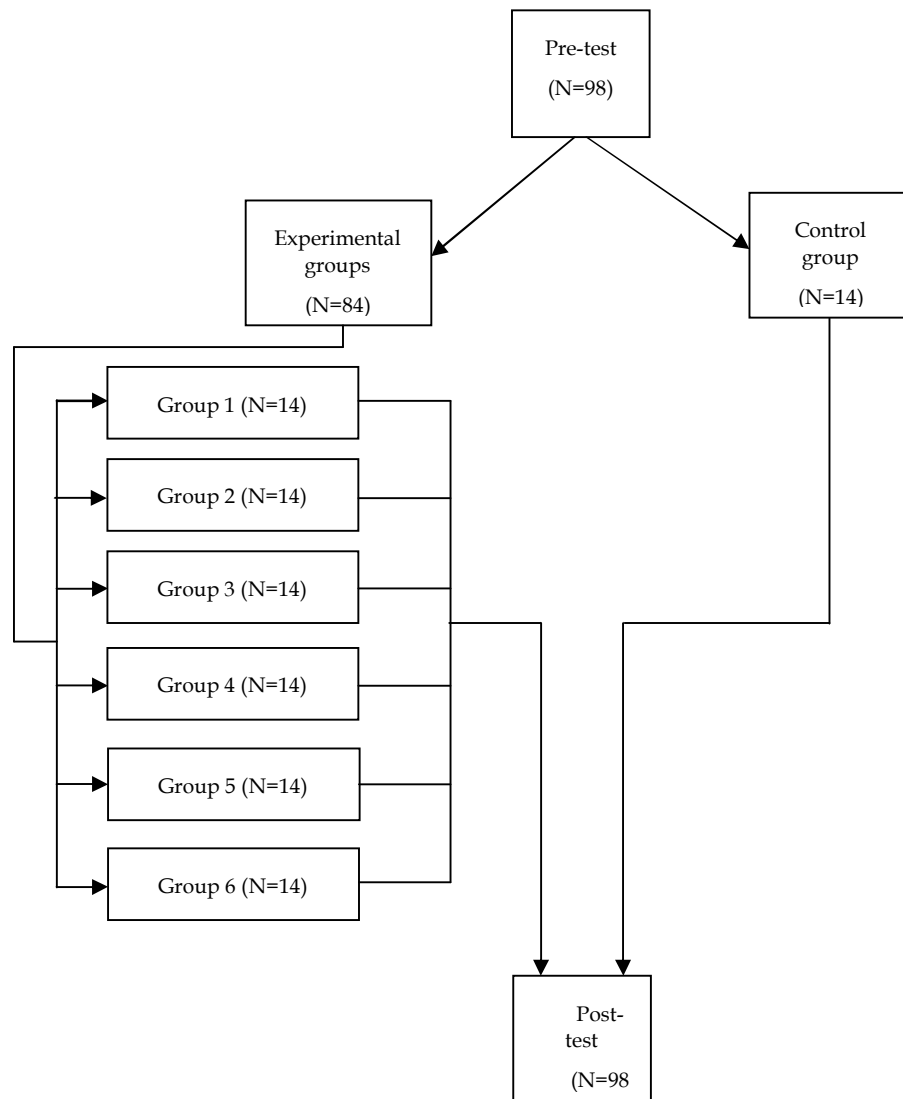


Figure 1. A schematic view of the study

Results

Descriptive Statistics

Descriptive statistics for pre-test and post-test scores of the groups are presented in Figure 2. Sub-groups of the study were formed based on their pre-test scores; mean pre-test scores of all sub-groups were equal with slightly different standard deviations. When post-test scores are considered, it is clear that all groups performed

better in the post-test. However, the increase in the mean score of the control group was very low; it only increased 3.78 points ($M_1=33.29$, $M_2=37.07$). Since the control group did not receive any intervention related to this topic in the classroom, this small difference may be explained by the practice effect, or students might have studied during the time period between pre-test and post-test administration.

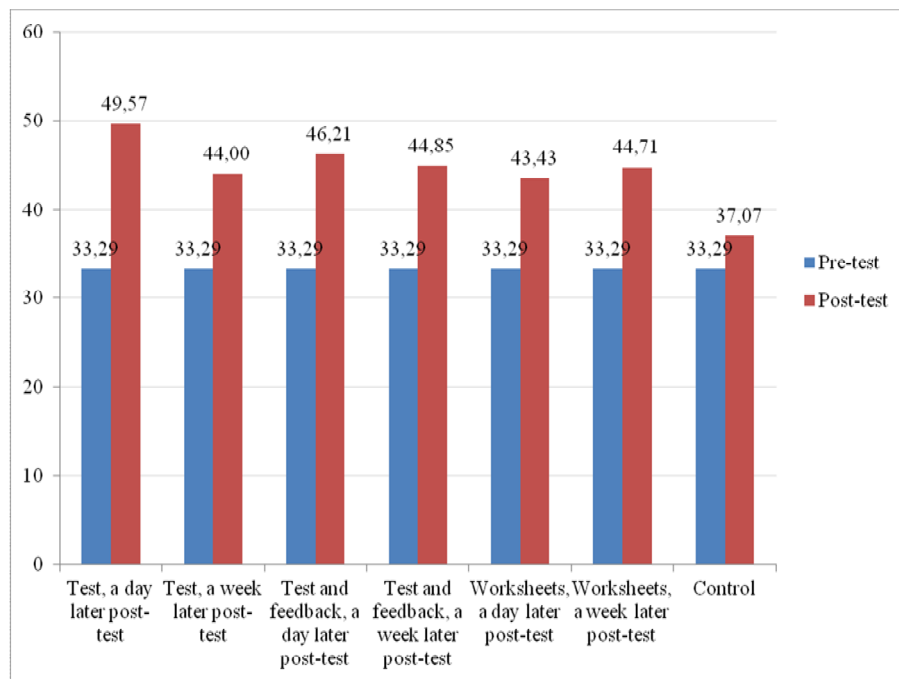


Figure 2. Pre-test & post-test scores of the sub-groups.

Wilcoxon Signed Rank Test Results

Descriptive statistics results (see Figure 2) showed that there is an increase in test scores of all sub-groups from pre-test to post-test. In order to investigate whether these increases were statistically significant, the Wilcoxon Signed Rank Test was used. The results of the test were presented in Table 2.

Table 2
Wilcoxon Signed Rank Test results

Group	Pretest - Posttest	
	Z	Asymp. Sig. (2-tailed)
Test, a day later post-test	-3.297 ^a	0.001
Test, a week later post-test	-3.112 ^a	0.002
Test and feedback, a day later post-test	-2.860 ^a	0.004
Test and feedback, a week later post-test	-3.298 ^a	0.001
Worksheets, a day later post-test	-2.732 ^a	0.006
Worksheets, a week later post-test	-3.235 ^a	0.001
Control	-1.855 ^a	0.064

a. Based on positive ranks.

The Wilcoxon Signed Rank Test revealed a statistically significant increase in mean scores of all sub-groups that participated in interventions. However, for the control group, who did not receive any intervention, the increase in mean score was not statistically significant ($p > 0.05$).

The effect size for this test can be calculated by dividing Z value by the square root of N, where N is the number of observations over the two time points ($14 \times 2 = 28$ for each sub-group), (Pallant, 2007). Effect size values for each practice sub-group were calculated and found to indicate a large effect (see Table 3). According to Cohen (1988), ≥ 0.1 indicates small effect, ≥ 0.3 indicates medium effect and ≥ 0.5 indicates a large effect.

Table 3
Effect size values

Group	Effect size (r)
Test, a day later post-test	0.62
Test, a week later post-test	0.59
Test and feedback, a day later post-test	0.54
Test and feedback, a week later post-test	0.62
Worksheets, a day later post-test	0.52
Worksheets, a week later post-test	0.61

ANCOVA Results

A one-way between-groups analysis of covariance was conducted to compare effectiveness of the three different interventions designed to increase students' test scores with the effect of time. The independent variables were the type of intervention (test, test and feedback, worksheets) and the time, and the dependent variable consisted of post-test scores. Students' pre-test scores were used as the covariate of the analysis.

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. After adjusting for pre-test scores, there was no significant difference between the practice groups on post-test scores, $F(5, 77)=0.80$, $p=0.55$, partial eta squared=0.05. There was a strong relationship between the pre-test and post-test scores, as indicated by a partial eta squared value of 0.59. Since there were no differences found between any two of the groups, no follow-up analysis was conducted.

Discussion and Conclusion

In this study, six practice sub-groups and a control group were formed according to their pre-test scores, and the results of the analyses showed that with the administration of tests and worksheets significant differences arose between practice sub-groups' mean pre-test and post-test scores. In the control group, which did not receive any intervention, there was not a significant difference between mean pre-test and post-test scores. However, although this difference was not statistically significant, the mean post-test score of the control group was slightly higher than its mean pre-test score. This small increase in the mean score of the group might have resulted from the re-exposure of the testing material, which is one of the explanations of the testing effect. The use of a pre-test may create a practice effect that can affect the results; practice on the pre-test by itself may be responsible for the improvement (Fraenkel & Wallen, 2006).

The significant difference between mean pre-test and post-test scores of each practice sub-group indicated that practicing and testing helped with the retention of previously learned information. When the literature was reviewed, many studies were found with similar results, suggesting that practicing and testing helped with the retention of previously learned information (Butler & Roediger, 2007; Agarwal, Karpicke, Kang, Roediger & McDermott, 2007; Roediger & Karpicke, 2006a, 2006b; Wheeler, Ewers & Buonanno, 2003; Chang, Yeh & Barufaldi, 2010; McDaniel, Anderson, Derbish, & Morrisette, 2007; McDaniel, Roediger & McDermott, 2007). As a result of educational reform, using methods that aim to evaluate process in addition to traditional assessment and measurement methods is inevitable. But for assessment and measurement, tests are frequently preferred in education since they are time-saving and easy to administer and evaluate.

According to Dempster (as cited in Roediger & Karpicke, 2006b), there are two possible explanations for the positive effects of testing on learning: (1) the testing effect may be a result of additional exposure to learning material during a test and (2) tests may enhance learning via retrieval processes that work on memory. When similar studies were investigated, it was discovered that the testing effect was mostly studied in the field of psychology, in laboratory settings in which administrations occur in a short time period, rather than in the field of education, in which studies generally require a longer time period. In the present study, while investigating the testing effect, test administrations and interventions were done in a classroom setting and, instead of studying the current topic, a previously taught topic was used to study retention of the previously learned information. With this aspect of the study, it can be claimed that testing is effective even after some time passes after learning information. Therefore, in this study, it is possible that the testing effect resulted from tests' enhancing power on learning via retrieval processes rather than additional exposure to learning material via tests.

When the mean pre-test and post-test scores of the practice sub-groups (G5 and G6), to which worksheets were administered, were compared, a significant difference was found. This result indicated that worksheets or a summary of the lecture notes also help with retention of previously learned information. However, Butler and Roediger (2007) conducted a study to determine different types of lecture materials' effect on retention of previously learned information and the results indicated that short-answer exams were superior to multiple-choice tests and worksheets on retention of previously learned information. The difference between the results of that study and the present study may have resulted from differences in learning and study styles of the participants or differences in educational policies of the two countries. In the country in which present study was conducted, the re-studying method was preferred by most of the students in examination periods. Most of the students use this technique to get ready for their examinations. It is thought that this situation affects the results. Moreover, although the difference between the mean post-test scores of the practice sub-groups was not statistically significant, when Figure 2 is examined, it can be seen that the practice sub-groups to which tests were administered performed better on the post-test than the practice groups to which worksheets were administered. Therefore, it can be said that tests may be superior to work sheets on retention of previously learned information. In the study of Roediger and Karpicke (2006a), in which the effects of testing and re-studying on remembering words in previously read paragraphs were investigated, post-tests were administered after different time periods (5 min, 2 days and a week), and it was found that tests were superior to re-studying for remembering previously learned information.

In this study, mean post-test scores of the practice sub-groups were compared in order to investigate whether the effects of different interventions differed. It was seen that mean post-test scores of the sub-groups were different from each other, and practice type had an effect on retention of the previously learned information, but these differences were not statistically significant. Moreover, when the effect of time

passed between pre-test and post-test administration was investigated, it was found that among the practice sub-groups (test, test and feedback, worksheets), administration of post-test a day later or a week later did not create a statistically significant difference. However, in test and test and feedback groups (G1 to G4), mean post-test scores of the groups to which post-test was administered a week later were lower than the groups to which post-test was administered a day later. Similarly, Roediger and Karpicke (2006a) found that as the time period between pre-test and post-test administration was strung out, retention of the previously learned information decreased.

Another finding of this study was that feedback given immediately after the test administration did not create a significant difference in retention of the previously learned information. When similar studies were investigated, in accordance with this study's results, Butler and Roediger (2007) also concluded that regardless of the test type, feedback does not have a significant effect on retention of previously learned information. On the other hand, there are also studies with findings supporting feedback as being effective on recall (McDaniel, Roediger, & McDermott, 2007). Moreover, the time period between feedback and test administration, and the allocated time for giving feedback, also influence the effectiveness of feedback on retention of previously learned information. Feedback given within a short time period may not give students enough time to process given information. For this reason, sufficient time should be allocated for giving feedback. In this study, feedback was found to be ineffective on retention of previously learned information; this result might be explained with the short time allocated for giving feedback and the short time period between giving feedback and test administration.

In summary, the results of this study showed that exposing students to supportive practices has a positive effect on retention of previously learned information regardless of the type of the practice. Specifically, tests, which educational professionals frequently use to assess their students' learning, should be used to support teaching and learning processes and not just to determine the level of learning.

Implications

The results of the present study, as well as the number of other studies investigating the testing effect, have important implications for classroom practice. That is, since much research supports the claim that testing has an important effect on students' retention of previously learned information, it therefore should be used to improve classroom practice and support teaching and learning processes. Test use should be encouraged in educational settings not only for evaluation purposes but also for learning purposes. However, they should not be used as alternatives to lecture notes, but as supporting materials. Future research needs to investigate the effect of feedback in a more detailed way; for instance, feedback may not be effective when given immediately after testing. Moreover, the time period between testing and post-test might be lengthened, and also the effect of other test types on retention of previously learned information can be investigated.

References

- Agarwal, P.K., Karpicke, J.D., Kang, S.H.K., Roediger III, H.L. & McDermott, K.B. (2007). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22, 861-876.
- Bangert-Drowns, R.L., Kulik, J.A., & Kulik, C.L.C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, 85, 89-99.
- Butler, A.C. & Roediger, III, H.L. (2007). Testing improves long term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4), 514-527.
- Carpenter S.K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563-1569.
- Carpenter S.K. & DeLosh, E.L. (2006). Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of testing effect. *Memory & Cognition*, 34(2), 268-276.
- Chang, C.Y., Yeh, T.K. & Barufaldi, J.P. (2010). The positive and negative effects of science concept tests on student conceptual understanding. *International Journal of Science Education*, 32(2), 265-282.
- Cohen, J. W. (1988). *Statistical power analysis for the behavioural sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Ass.
- Creswell, J.W. (2005). *Educational research planning, conducting, and evaluating quantitative and qualitative research*. (2nd ed.). Pearson Education, New Jersey.
- Cull, W.L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14; 215-235.
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education*. (6th ed.). McGraw-Hill, Inc.
- Johnson, C.I. & Mayer, R., E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, 101(3), 621-629.
- Kang, S.H.K., McDermott, K.B. & Roediger, III, H.L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19(4/5), 528-558.
- Karpicke, J.D. & Zaromb, F.M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, 62, 227-239.
- Leeming, F.C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, 29(3), 210-212.
- McDaniel, M.A. Anderson, J.L. Derbish, M.H. & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of the Cognitive Psychology*, 19, 4/5; 494-513.

- McDaniel, M.A, Roediger III, H.L. & McDermott, K.B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14(2), 200-206.
- Mintzes, J. J., Wandersee, J. H. & Novak, J. D. (2001). Assessing understanding in biology. *Journal of Biological Education*, 35(3);118-124.
- Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows* (3rd edn). Maidenhead, UK & New York: Open University Press.
- Roediger III, H.L. & Karpicke, J.D. (2006). Test-enhanced learning, *Psychological Science*, 17(3); 249-255.
- Roediger III, H.L. & Karpicke, J.D. (2006). The power of testing memory, *Perspectives on Psychological Science*, 1(3), 181-210.
- Thomas, A.K. & McDaniel, M.A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, 35(4), 668-678.
- Wartenweiler, D. (2011). Testing effect for visual-symbolic material: enhancing the learning of Filipino children of low socio-economic status in the public school system. *The International Journal of Research and Review*, 6(1), 74-93.
- Wheeler M., Ewers, M. & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11(6), 571-580.

Sınıflarda Test Etkisinin Değerlendirilmesi: Öğrenilmiş Bilgilerin Geri Çağırılmasında Etkili Bir Yol

Atf:

- Atabek Yiğit, E., Balkan Kıyıcı, F. & Çetinkaya, G. (2014). Evaluating the testing effect in the classroom: An effective way to retrieve learned information, *Eurasian Journal of Educational Research*, 54, 99-116.

Özet

Problem durumu: Öğretim uygulamalarında önemli bir adım olan değerlendirme genellikle öğrencilerin ne bildiklerini ya da ne kadar öğrendiklerini ölçmek amacıyla yapılan bir işlem olarak düşünülmektedir. Ölçme işlemini gerçekleştirmek için farklı yollar kullanılabilir ve testler bu yollar arasında en önemlisi ve en yaygın olarak kullanılanıdır. Testler kullanılarak gerçekleştirilen ölçme ve değerlendirme işlemleri genelde uzun zaman gerektirmediği için tercih edilen bir yöntemdir. Testlerin hazırlanma aşaması zaman alıcı bir ölçme yöntemi olmasına rağmen, kalabalık gruplar için kolaylıkla uygulanabilir ve objektif olarak puanlaması yapılabildiği için yaygın olarak kullanılmaktadır. Ayrıca öğrencilerde ölçme aracı olarak testleri tercih etmektedirler. Özellikle doğru cevabı bulma şansına sahip

olabilme, öğrenciler için testleri öncelikli olarak tercih edilebilir yapmaktadır. Yapılan çok sayıda araştırmaya göre; testlerin uygulanması sırasında zihinde gerçekleşen bir takım zihinsel aktiviteler yardımıyla öğrenciler eski bilgilerini hatırlayabilir ya da yeni bir öğrenme işlemi gerçekleştirebilirler. Bu eski bilgileri hatırlama ve öğrenme işleminin gerçekleşmesi işlemine kısaca test etkisi ismi verilmektedir. Test uygulamalarının sağladığı avantajlarının yanında bazı dezavantajları da olmasına rağmen, öğrencilerin sıklıkla karşı karşıya kaldıkları testler, test etkisi dolayısıyla sınıflarda bir öğrenme materyali olarak kullanılabilir. Bu sebepten bu çalışmada, sınıflarda test etkisi araştırılmaya çalışılmıştır.

Araştırmanın amacı: Bu araştırmanın amacı sınıflarda test etkisinin araştırılmasıdır. Test etkisini belirleyebilmek amacıyla çoktan seçmeli test, eşleştirme testi ve konuyu özetleyen bir çalışma yaprağı (tekrar çalışma grubu için) kullanılmıştır. Araştırma grubuna ön test olarak 100 kısa sorudan oluşan bir ön test uygulanmış ve öğrenciler ön test sonucuna göre gruplandırılmışlardır. Ön test sonuçlarına göre öğrenciler her bir grubun ortalama puanı aynı olacak şekilde ayarlanarak yedi gruba (altı grup deney grubu bir grup kontrol grubunu oluşturacak şekilde) ayrılmışlardır. Ön test uygulamasından sonra, 4 deney grubuna müdahale testi uygulanmış ve bu gruplardan 2 gruba test sonrasında dönüt verilmiştir, deney gruplarından 2 gruba ise müdahale testi uygulanmamış ve çalışma kağıdı verilmiştir. Kontrol grubuna ise herhangi bir işlem uygulanmamıştır. Öğrencilerin son öğrenmelerini belirlemek üzere ön test olarak kullanılan test son test olarak da uygulanmıştır. Deney gruplarından üç gruba son test bir gün sonra uygulanırken, diğer 3 deney grubuna ve kontrol grubuna son test bir hafta sonra uygulanmıştır. Sonuç olarak; öğrencilerin öğrendikleri bilgileri hatırlama düzeylerinde test etkisi, dönütün hatırlama üzerine etkisi ve tekrar çalışmanın hatırlama üzerine etkisi araştırılmıştır.

Araştırmanın Yöntemi: Bu çalışmada Eğitim Fakültesinde yer alan Genel Kimya dersinde öğrenilen bileşiklerin isimlendirilmesi konusu ele alınarak test etkisi belirlenmeye çalışılmıştır. Bu konu öğrenciler için anlaşılması zor ve kolay unutulmuş bir konudur. Araştırmanın örneğini bu dersi alan fen bilgisi öğretmen adayları oluşturmakta olup, araştırmaya katılmaya gönüllü olduğunu belirten 98 fen bilgisi öğretmen adayı oluşturulmuştur. Araştırmada test etkisini belirlemek üzere ön-test, son-test kontrol gruplu araştırma deseni kullanılmıştır.

Araştırmanın Bulguları: Araştırma sonucunda ön test puanları eşit olan altı deney grubu ve kontrol grubuna uygulanan son test puanlarından elde edilen veriler istatistiksel olarak analiz edilmiştir. Araştırma sonuçları, uygulama tipinin ne olduğu önemli olmaksızın öğrenilmiş bilgilerin hatırlanmasında, destek etkinliklerinin olumlu bir etkiye sahip olduğunu göstermektedir. Bu çalışmada testlerin geri çağırma sürecine yardımcı olmasıyla test etkisi sonucundan söz etmek mümkündür. Aynı zamanda araştırmanın sonuçlarından bir diğerine göre; çalışma yaprakları da öğrenilmiş bilgilerin hatırlanmasında öğrencilere yardımcı olmaktadır. Bu sonuç literatürdeki diğer çalışmalarla benzer bir sonuca işaret etmemektedir. Ancak çalışmanın yapıldığı ülkede sınav dönemlerinde öğrencilerin sınava hazırlık için çoğunlukla bu metodu tercih ediyor olmasının sonucun bu şekilde çıkmasını etkilediği düşünülmektedir. Farklı deney gruplarındaki ön test ve son test

uygulamaları arasında zaman farkı dikkate alınarak yapılan analizler sonucunda ise son testlerin bir gün sonra veya bir hafta sonra uygulanmasının herhangi bir önemi olmaksızın, istatistiksel olarak anlamlı bir fark oluşturmadığı tespit edilmiştir. Bu çalışmanın sonuçlarından bir diğeri ise; öğrencilere testlerin arkasından dönüt verilmesinin öğrenilen bilgilerin hatırlanmasında gruplar arasında anlamlı bir farklılık oluşturmadığıdır. Bu sonuçlardan hareketle özellikle eğitim uzmanları tarafından öğrenci öğrenmelerini belirlemek amacıyla sıklıkla kullanılan testlerin, öğrenme seviyesine belirlemenin yanı sıra, öğrenme öğretme süreçlerini desteklemek amacıyla da kullanılabileceğini söylemek söz konusudur.

Araştırmanın Sonuç ve Önerileri: Araştırma sonuçları göstermektedir ki; öğrenciler testler ve çalışma yaprakları ile karşı karşıya kaldığında bu uygulamalar öğrenilmiş bilgilerin hatırlanmasında öğrencilere yardımcı olmaktadır. Araştırma sonuçları sınıf etkinlikleri için önemli uygulamalar geliştirmeye yardımcı olabilecek niteliktedir. İlgili literatür incelendiğinde; yapılan birçok araştırma da test uygulamalarının öğrenci performansı ve daha önce öğrenilen bilgileri hatırlama üzerinde önemli etkisi olduğunu göstermektedir. Dolayısıyla test uygulamaları, sınıf içi öğrenme uygulamalarını geliştirmek ve öğrenme öğretme süreçlerini desteklemek amacıyla kullanılmalıdır. İleriki araştırmalar için dönüt etkisinin daha ayrıntılı olarak araştırılması önerilebilir, örnek olarak dönüt test uygulamasının hemen arkasından verilmeyip, daha sonra verildiği araştırmalar planlanarak etkisi değerlendirilebilir.

Anahtar sözcükler: Test etkisi, dönüt, hatırlama, geri çağırma, fen eğitimi