

**AN EXPERIMENTAL STUDY**  
**ON**  
**THE EFFECTS OF FREQUENT TESTING**

**A THESIS**  
**SUBMITTED TO THE FACULTY OF LETTERS**  
**OF BILKENT UNIVERSITY**  
**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS**  
**FOR THE DEGREE OF MASTER OF ARTS**  
**IN THE TEACHING OF ENGLISH AS A FOREIGN LANGUAGE**

21533

**BY**  
**MELTEM TARHAN**  
**AUGUST, 1992**

**T.C. YÜKSEKÖĞRETİM KURULU**  
**DOKÜMANTASYON MERKEZİ**

2533

BILKENT UNIVERSITY  
INSTITUTE OF ECONOMICS AND SOCIAL SCIENCES  
MA THESIS EXAMINATION RESULT FORM

August, 1992

The examining committee appointed by the  
Institute of Economics and Social Sciences for the  
thesis examination of the MA TEFL student

MELTEM TARHAN

---

has read the thesis of the student.  
The committee has decided that the thesis  
of the student is satisfactory.

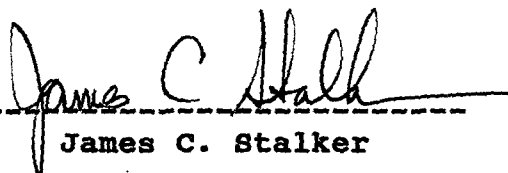
Thesis title : An Experimental Study on the  
Effects of Frequent Testing.

Thesis Advisor : Dr. James C. Stalker  
Bilkent University, MA TEFL Program

Committee Members: Dr. Lionel Kaufman  
Bilkent University, MA TEFL Program

Dr. Eileen Walter  
Bilkent University, MA TEFL Program

We certify that we have read this thesis and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts.

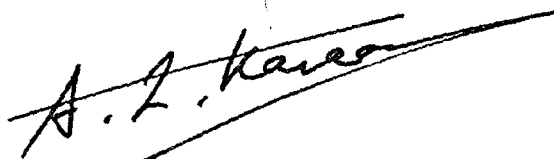
  
-----  
James C. Stalker

  
-----  
Lionel Kaufman

  
-----  
Eileen Walter

Approved for the  
Institute of Economics and Social Sciences

-----  
Ali Karaosmanoglu  
Director  
Institute of Economics and Social Sciences





To my husband Halil  
for his patience ...

## TABLE OF CONTENTS

LIST OF TABLES	vii
1.0 INTRODUCTION	
1.1 Background and Goals of the study	1
1.1.1 Background of the study	1
1.1.2 Goals of the study	5
1.2 Statement of Research Topic	6
1.3 Hypothesis	7
1.3.1 Null Hypothesis	7
1.3.2 Experimental Hypothesis	7
1.4 Overview of Methodology	8
1.5 Overview of Analytical Procedures	9
1.6 Organisation of Thesis	10
2.0 REVIEW OF LITERATURE	
2.1 Introduction	11
2.2 The Effects of Frequent Testing	11
2.2.1 Research Supporting Frequent Testing	11
2.2.2 Research Opposing Frequent Testing	13
2.3 Discrete-point Tests	16
2.4 Cloze Tests	20
2.5 Practice Effect	24
2.6 Summary	26
3.0 METHODOLOGY	
3.1 Introduction	28
3.2 Subjects	31
3.3 Materials	33
3.3.1 Pre-test/Post-test	33
3.3.2 Quizzes	34

## LIST OF TABLES

TABLE 4.1	Means and Standard Deviations of Pre- and Post-test Scores for Each Group	42
TABLE 4.2	T-values of the Pre- and Post-test Scores for Each Group	43
TABLE 4.3	Total Gain Scores Between Pre- and Post-test	44
TABLE 4.4	One-way Anova Analysis for the Gain Scores of Each Group	44
TABLE 4.5	Scheffe Test Analysis for Gain Scores	45
TABLE 4.6	Gain Scores for Structural Transformation Group on Cloze and Structural Transformation Sections on the Post-test.	46
TABLE 4.7	Gain Scores for Cloze Group on Cloze and Structural Transformation Sections on the Post-test.	47
TABLE 4.8	Gain Scores for Control Group on Cloze and Structural Transformation Sections on the Post-test.	47
TABLE 4.9	The Mean and the Standard Deviations of All Subjects in the Cloze and the Structural Transformation Sections	69
TABLE 4.10	One-way Anova Analysis for the Gain Scores on the Cloze Section of the Post-test.	48

3.3.3	Teacher Questionnaire	36
3.3.4	Student Questionnaire	37
3.4	Procedures for Data Collection	37
3.5	Variables	38
3.5.1	Independent Variables	38
3.5.2	Dependent Variables	39
3.6	Analytical Procedures	39
4.0	DATA ANALYSIS	
4.1	Introduction	41
4.2	Data Analysis	42
4.2.1	Hypothesis 1	42
4.2.2	Hypothesis 2	45
4.2.3	Hypothesis 3	50
4.3	Interpretation and Discussion	53
5.0	CONCLUSIONS	
5.1	Summary of the Study	56
5.2	Assessment of the Study	59
5.3	Pedagogical Implications	60
5.4	Implications for Future Research	61
	BIBLIOGRAPHY	64
	APPENDICES	
	APPENDIX A PRE-TEST	67
	APPENDIX B POST-TEST	71
	APPENDIX C EXAMPLES OF QUIZZES	75
	APPENDIX D TEACHERS' QUESTIONNAIRE	79
	APPENDIX E STUDENTS' QUESTIONNAIRE	80

TABLE 4.11	One-way Anova Analysis for the Gain Scores on the Structural Transformation Section of the Post-test.	49
TABLE 4.12	Scheffe Analysis for the Gain Scores on the Structural Transformation Section on the Post-test	50
TABLE 4.13	Descriptions of Items in the Questionnaire and Their Mean Scores	51
TABLE 4.14	The Students' Post-test and Attitude Scores	52
TABLE 5.1	Mean Scores of Progress Tests for All Intermediate BUSEL Classes.	60

## ACKNOWLEDGEMENTS

I would like thank Dr. James Stalker, my thesis advisor, without whose invaluable guidance, encouragement and support this thesis would have never been completed.

I am also grateful to Dr. Lionel Kaufman for his patient guidance on research techniques and statistical computations during the writing of this thesis.

I also thank Dr. Eileen Walter for her support in carrying out this thesis.

I must express my special thanks to my colleagues Nalan Eren, Halis Cetin and Sally Kaufman, whose classes were selected as the experimental and the control groups, for their co-operation; and Ali Nihat Eken for his help while writing the quizzes.

I must also thank my wonderful friends Sibel and Isik Okan and Dilek and Hakan Birbil who have encouraged me even at my most desperate times, and who also lent me their printers which made life easier for me.

Finally, I must express my greatest debt to my husband, my parents, my sister and my brother for making me feel their support all the time while writing this thesis.

=

**AN EXPERIMENTAL STUDY ON THE EFFECTS  
OF FREQUENT TESTING**

**Abstract**

Testing has been researched extensively because it is considered to be an important tool for providing students' with better learning. Thus, this study was done to find out whether frequent testing would increase students' scores on a post-test, whether there would be a practice effect on students' scores on particular types of questions which they had practiced, and finally, what students' attitudes toward frequent testing were.

The effect of frequent testing was investigated through a t-test analysis of the pre- and the post-test scores to determine whether frequent testing of the students within the treatment period, between the pre- and the post-tests, helped students get higher scores on the post-test. The findings showed that giving frequent quizzes to students did not have the effect of increasing students' success on a post-test, despite what had been hypothesized at the beginning of the study. However, the analysis of the gain scores of the students in the three groups, which was handled as a secondary issue and analysed through one-way Anova, and Scheffe, a post hoc test, indicated that the experimental group which received discrete-point quizzes performed better than the other experimental group, which received integrative type quizzes, and the control group,

which received no quizzes. Therefore, this performance of the discrete-point group, which was significantly better than the other two groups in terms of their gain scores, may suggest that discrete-point quizzes provided an advantage to the students who took this type of quiz over the ones who did not.

The study on practice effect showed that the structural transformation group performed significantly better than the cloze group, on the type of questions they had practiced, however the cloze group could not perform as significantly as the structural transformation group. This was determined through one-way Anova and Scheffe analyses, and as a result, the null hypothesis on the practice effect could only be partially rejected.

Finally, the study found no significant relationship between students' attitudes toward frequent testing and their scores on the post-test. A correlational analysis, PPMC, was carried out to correlate the students' over-all scores on the post-test with their answers to a questionnaire.

## CHAPTER 1

### INTRODUCTION

#### 1.1 BACKGROUND AND GOALS OF THE STUDY

##### 1. 1. 1 Background of the study

Testing is believed to be a very useful tool to increase students' success (Bloom et al., 1971; Ebel, 1980; Natriello & Dornbush, 1984; Peckham & Roe, 1977). It is also accepted as one of the most important issues in education. Oller (1979) describes language tests as devices which help the teacher to evaluate how much the students have learned and which also help the teacher to locate the students' weaknesses. Freilich (1989) states that frequent quizzing may help the students "improve" the quality of their knowledge by guiding them in recognizing their weaknesses or forcing them to "study more regularly" (p.220). Madsen (1983) proposes that the diagnostic characteristics of tests help the teacher to confirm which subjects are mastered by each individual and which subjects still need to be reviewed. They also help students to focus on the areas of emphasis on which they need to concentrate their studies. Crooks (1988) says that classroom evaluation guides the students to "what is important to learn" (p.467) and affects their motivation. However, he further adds that the greater use of peer review should be encouraged as well. Madsen (1983) also adds that with the results of the tests, a feeling of accomplishment will motivate

the students even more. Furthermore, the teacher can share the same feeling of accomplishment because the grades tell the teacher if she is effective or if the lessons are taught at the right level. Therefore, the grades that the students get from the tests have the function of motivating the teacher as well.

Bilkent University School of English Language (BUSEL) teachers tend to give quizzes as well as summative tests for some or all of the reasons explained above. BUSEL teachers also follow the tendency to test grammar despite the views of Oller (1979) and Carroll (1982) who find the reasons for evaluating students' improvement by grammar tests obscure, since recent innovations in ELT mostly focus on skills such as reading and writing rather than grammar. Furthermore, BUSEL teachers prefer to use discrete-point tests, although Oller (1979) says discrete-point tests are difficult to prepare.

Despite the general tendency to shift emphasis from the traditional method of testing the knowledge of language form to the ability to use the language to communicate (Bozok & Hughes, 1984), the general situation in BUSEL is in opposition to this tendency, most particularly in the area of teacher-made classroom examinations. Instead of using integrative skill tests which would be a better reflection of their teaching approaches, teachers in BUSEL tend to provide the students with discrete-

point grammar tests which focus on one point of grammar at a time. Hughes (1989) says that there are some institutions which are in favor of using communicative methods, but still have a grammar component in their examinations. He further explains that:

Whether or not grammar has an important place in an institution's teaching, it has to be accepted that grammatical ability, or the lack of it, sets limits to what can be achieved in the ways of skills performance. The successful writing assignments, for example, must depend to some extent on command of more than the most elementary grammatical structures. It would seem to follow from this that in order to place students in the most appropriate class for the development of such skill, knowledge of a student's grammatical ability would be very useful information. There appears to be room for a grammar component in at least some placement tests. (p. 142)

Hughes' justification of the use of a grammar component in tests can easily be applied to BUSEL, where the teachers tend to use discrete-point tests in their classroom examinations.

Because the scores the students get from these quizzes are not counted on their final grades, the reason that BUSEL teachers give quizzes with discrete-point items cannot be to assess the students just for grades. Their aim is more likely to evaluate their students' knowledge and locate where their weaknesses are. Therefore, the aim of this study is to explore the theoretical justification for initiating a continual assessment system in BUSEL

where the quizzes would be given to improve the students' success in learning English.

Frequent testing is also considered to be beneficial by A. Harris (1975) and D. Harris (1969). A. Harris (1975) says every conscientious teacher has to teach his/her students as much as possible, and therefore monthly and weekly quizzes should be used in fulfilling that obligation. D. Harris (1969), furthermore, adds another dimension to the positive effects of frequent testing when he says that frequently repeated test items that are similar to those on the post-test or final examination will likely result in an increase in the final scores. Therefore BUSEL teachers might also want to create some familiarity with the test styles by giving quizzes with items similar to those in the final examination. As D. Harris (1969) explains, the practice effect must be taken into consideration:

Generally speaking, the more often we perform an operation, the more proficient we become at it...Thus we may expect subjects who are repeating a test (whether with the same or a parallel form) to score somewhat higher than they did the first time, even if their knowledge of the subject being tested has not itself increased. Test users must therefore make allowance for "practice effect" when evaluating scores on "progress" or "exit" tests. (p. 130)

Therefore, another focus of this study was to measure the effects of periodic quizzes and determine to what extent the effect of practising the test items helped the students get higher grades

on the post-test and what the attitudes of students toward frequent testing were.

### 1.1.2 Goals of the Study

Testing can be very beneficial. Slavin (1986) gives five reasons why students are evaluated by tests and quizzes:

1. Incentives to increase student effort
2. Feedback to students
3. Feedback to teachers
4. Information to parents
5. Information for selection (p. 541)

In the current study, the researcher hypothesized that frequent testing in BUSEL would meet at least the first three reasons set by Slavin because, through in-class evaluation through frequent quizzes, the students as well as the teachers would have feedback to offer positive guidance. Corrections in class would serve a diagnostic function by which the students would know where to improve. Preparing the quizzes with items similar to those on the post-test could also increase scores. This is a direct result of the practice effect caused by the similarity between the items in the quizzes and the post-test.

Therefore, the goal of this study was to determine whether frequent discrete-point grammar tests and cloze tests would increase students' success on a post-test and whether the practice effect would cause a gain on the sections of the tests in which these type of questions were asked. A significant increase in the performance on the post-test in the experimental groups would suggest that

initiating a system of frequent assessment through quizzes would be beneficial in terms of test performance.

## **1.2 Statement of the Research Topic**

The purpose of this study was to find out if frequent testing increased students' success on a post-test. In addition, the relationship between practice effect and degree of success was studied. As a third point, students' scores on the post-test were correlated with a questionnaire gauging students' attitudes toward frequent testing.

The researcher expected to observe an increase on students' post-test scores as a result of the students' continual evaluation by weekly discrete-point grammar and cloze quizzes. The researcher also expected to find a relationship between the practice effect and the degree of success. A positive correlation between the students' scores on the post-test and the results of the questionnaire about students' attitudes toward frequent testing was also expected.

The terms used in this statement are as follows:

- \* **Discrete point tests:** Testing of one element at a time, item by item (Hughes, 1989).
- \* **Cloze test:** Prose passages from which the words are deleted randomly and the students rely on the context in order to supply the missing word. (Madsen, 1983)
- \* **Post test:** The midterm test prepared by the BUSEL

Testing Office which was given to the control and the experimental groups after the treatment.

\* **Practice effect:** The situation which occurs when the familiarity with the test types causes an increase in the scores when the knowledge of the subject being tested has not itself increased (D. Harris, 1969).

### **1.3 Hypothesis**

#### **1.3.1 Null Hypothesis**

Three related hypotheses are the focus of this study. The null versions of these are:

1. There is not a significant increase on students' post-test scores due to weekly discrete-point grammar and cloze quizzes.
2. There is no significant relationship between repeating the same type of questions to those on the post-test and gains on the sections in which those type of questions were asked.
3. There is no significant positive correlation between the students' scores on the post-test and the results of the questionnaire about students' attitudes toward frequent testing.

#### **1.3.2 Experimental Hypothesis**

The experimental versions of the three hypotheses are:

1. There is a significant increase on students' post-test scores from their pre-test scores due to weekly discrete-point grammar and cloze quizzes.
2. There is a significant relationship between

repeating the same kind of questions on the weekly quizzes as those on the post-test and gains on the sections in which those type of questions were asked.

3. There is a significant positive correlation between the students' scores on the post-test and their attitudes toward frequent testing as recorded in a questionnaire.

The variables which are the focus of the study are:

\* **Independent variables:** Number and type of quizzes

\* **Dependent variable :** Students' degree of success on a post-test test.

#### **1.4 Overview of the Methodology**

The experiment was conducted in 3 lower intermediate classes in Bilkent University School of English Language (BUSEL). In order to start the collection process, classes with similar proficiency levels were selected. To do this, the results of the elementary level second midterm (the last examination of winter semester) were examined and those classes which had standard deviations and averages very close to each other were chosen. A questionnaire was then given to the teachers of those classes to learn about the teachers' teaching strategies. It was necessary to find the teachers for the experimental groups who allocated time to quizzes, and to find teachers for the control group who normally did not give quizzes. The classes were selected after informing the teachers about the experiment and its protocol.

After the selection of the control and the two experimental classes, the data collection procedure started. There were 44 subjects, 21 female and 23 male BUSEL intermediate level students. Both the first and the second experimental groups received 10 quizzes with questions similar to those on the post-test. However, the types of quizzes were different. The first experimental group was given discrete-point items, and the second experimental group was given a cloze paragraph. The control group received no quizzes. After the treatment was completed, both experimental groups were given a questionnaire in order to measure their attitudes toward the treatment and to determine whether they felt quizzes helped them learn better. The post-test, which had both discrete-point items and a cloze paragraph as well as the other sections of standard mid-term examinations of BUSEL prepared by the Testing Unit, was then administered.

### **1.5 Overview of Analytical Procedures**

In order to consider the effects of frequent quizzing, practice effect and the attitude of the students' toward frequent testing, several statistical studies were done. The effect of frequent testing was determined through a series of t-test, one-way Anova and Scheffe analyses of the pre- and the post-test scores of the three groups.

The practice effect on the particular sections of the test was determined by using a one-way Anova

analysis. First, the gain scores of the students on these particular sections were found and then compared with each other.

The correlation between the students' attitude toward frequent testing and their post-test scores was done by using the Pearson Product Moment Correlation (PPMC).

### **1.6 Organisation of Thesis**

Of the five chapters in this thesis, the first chapter gives some background information on the positive effects of frequent testing, some discussions about discrete-point testing and the practice effect, states the hypothesis and gives a brief review of the methodology and analytical procedures. In Chapter 2, there is a detailed review of the research on the effects of frequent testing and the practice effect. Pros and cons of testing, examples of similar studies and a comparison of the different views about testing are given. Chapter 3 summarises the methodology of the data collection process and describes the statistical procedures used in the study. Chapter 4 displays the analysis of the data and interpretation of the results provided from the statistical analysis. Chapter 5 presents a summary of the study with conclusions and implications.

## **CHAPTER 2**

### **REVIEW OF LITERATURE**

#### **2.1 Introduction**

The effects of frequent testing have been rigorously studied from different aspects, since it is one of the most important components of education. Some researchers claim that there are positive effects of testing since testing gives both the teacher and the students a chance to perceive how much learning has been realised and tells them the points where weaknesses occur. Some researchers say that grading gives the students a feeling of achievement which motivates them to study and therefore increases their learning. There are, however, some researchers who oppose the idea of testing, claiming that tests increase anxiety so learning decreases.

#### **2.2 The Effects of Frequent Testing**

##### **2.2.1 Research Supporting Frequent Testing**

There are some researchers and scholars who support frequent testing. Peckham and Roe's (1977) findings on the positive effects of frequent testing, as researchers, support what Bloom et al. (1971) state. They agree that students achieve higher scores when more evaluations are done. They base this claim on the fact that frequent brief quizzes require students to pay continual attention on learning and studying rather than cram for the occasional examination. They also add that frequent

quizzes give students more timely feedback and provide reinforcement of the items to be learned. Halpin and Halpin (1982) accept quizzes as a tool that acts as both a guide and a prod to studying.

Martin and Srikameswaran (1974) investigated the impact of frequent quizzing upon examination performance. Fourteen sets of mastery-type quizzes, which were distributed over an academic year, supplemented normal testing. They found that the students who took the "mastery-type" quizzes, which supplemented the "normal" examinations, managed to get higher grades on the final examination than the ones who did not take the quizzes. In a similar study conducted by Gaynor and Millham (1976) in a psychology class, students who took weekly quizzes performed better on the final examination than the ones who took only a mid-term examination before the final examination. Natriello and Dornbush (1984) view the effectiveness of frequent testing as one of the six criteria that must be satisfied if the aim of the evaluations is to increase student effort. On a similar note, Ebel (1980) observes that students' learning can be best facilitated by determining how much learning has occurred and how successfully the student can cope with tasks that require learning. He adds that since grading gives us more meaningful and concrete indications of the degree of a student's success in learning, the best way to answer those questions is to test and grade the

students frequently.

On the other hand, Anderson and Faust (1973), in their book on educational psychology, claim that testing students for the sake of grading should not be the most important issue, because the most important function of evaluation should be providing a system of quality control. Therefore, these authors also claim that to cultivate the benefits and avoid the uncertainty of summative tests, the best thing to do is to give short tests immediately after the students have been taught the skills and knowledge.

### **2.2.2 Research Opposing Frequent Testing**

As well as the views supporting frequent testing, there are views claiming that frequent tests are not beneficial. Freilich (1989) found no effect of frequent testing on the success of the students on the final examination. Despite his claims supporting frequent testing due to its pedagogical implications rather than assessment, his research findings did not support his theoretical views. In his study, Freilich focused on the effects of frequent tests given at different intervals, and despite his hypothesis, he found that too often tests do not help students learn better. He further points to the importance of the amount of testing indicating that additional testing, beyond a certain point, apparently produces no discernible increase in the retention of information over the short term. Bangert-Drowns et al. (1986) obtained results from their study which

support Freilich's views. They found that the number of tests given to the high frequency group was not significantly correlated with the effect size. Therefore, he concludes that it is **not** true that a high frequency of testing results in higher performance by students.

Bangert-Drowns et al.'s (1986) study is related to the present research topic, because both the present study and the Bangert-Drowns et al. study aimed to determine the effects of frequent testing using similar methodologies. Nevertheless, there are some differences in their focus of measurement of the students' success. Instead of studying adjunct questions and mastery testing, in the present study the focus will be on grammar tested by discrete-point and cloze tests. There will be some differences in their research methodology, as well. In Bangert-Drowns et al.'s study there were several groups receiving quizzes at different frequencies, including the control group, as well. However, in this study there are only two groups receiving quizzes and there are no tests given to the control group during the experiment.

In the Bangert-Drowns et al.'s (1986) study, similar to Freilich's, no positive effects from frequent testing were found. Although it was found that the tested group performed better than the untested group, the reason for the better performance of the tested group did not seem to be due to the

higher frequency of tests, because a negative correlation was found between the frequency of tests and students' scores. Therefore, it was concluded that frequency of testing did not significantly influence academic achievement. For that reason, it remains unclear whether one can speak generally about the effects of frequent testing on the basis of evidence from the studies done by these researchers.

Bangert-Drowns et al. (1986) found no positive effect of frequent testing on students' scores, although their study had aimed to find a positive effect of frequent testing. They found testing is beneficial for the students but an excessive amount of testing creates a reverse effect on students; their scores decreased. The same question, how often should students be tested, was raised by Freilich in the evaluation section of his study. Since the present research also questions the effectiveness of frequent quizzes, if the findings of this study are parallel with that of Bangert-Drowns et al., then the implications of it should be considered very carefully.

Besides these researchers whose findings support negative views toward frequent testing, there are also some researchers or scholars who display a very strong opposition toward tests. Noll (1939) started his complaints about tests at a relatively early stage and his views obviously found supporters from scholars like Davies (1984). He argued that although

learners have a chance to demonstrate their knowledge by performing satisfactorily on tests, the tests themselves are a burden on the teaching, programs, materials, textbooks and syllabuses.

### **2.3 Discrete-point Tests**

In the history of testing, there have been many changes in testing techniques which are parallel with the changes in teaching methods. Farhady (1983), Heaton (1989), and Madsen (1983) classified testing approaches as evolving through three stages: the intuitive, the scientific and the communicative stages. Madsen explained the connection between these stages and the testing techniques as follows: the intuitive stage is associated with the grammar-translation approach where the examinations were mostly subjective. That is, techniques such as translation, precis writing, and dictation used in the examinations were scored on a holistic basis and were designed by teachers who assumed the tests were adequate instruments. The scientific stage occurred during the period when objective tests became predominant and the validity and the reliability of the tests became matters of concern. The communicative stage, the one we are currently in, emphasizes the use of language rather than the form of language. Therefore, in the communicative tests, the competency in the use of the language is tested rather than the knowledge about the form of the language.

The scientific stage, which Heaton (1989) associates with "the structuralist approach", developed discrete-point tests, because the essential goal was to test one thing at a time. Since validity and reliability of the tests gained importance at this stage, the reason that this type of test was recommended was a statistical necessity and not pedagogical one, because the only way to measure validity and reliability accurately is to test one thing at a time. Any item or test that allows two or three or more operations to happen at once cannot be handled by statistical measures, therefore cannot be accurately assessed for validity and reliability. Therefore, mastery of items or of the skills isolated from any context, demonstrated through multiple-choice examinations for the sake of objectivity, is the aim of this approach.

Discrete-point tests, however, have recently received a lot of criticism because of the structuralist view with which it is associated. Since the structuralist view has been regarded as ineffective when compared with more recent teaching techniques, discrete-point tests have been disfavored as well. Although Carroll was the first one to make the distinction between discrete-point and integrative tests, Oller is known as the warrior against discrete-point tests. For example, Oller (1979) criticizes discrete-point tests very sharply: "Discrete-point tests are notoriously

ineffective" and "Discrete-point methods don't work" (p. 211).

However, there are also some other researchers who oppose Oller because of his very negative criticisms on discrete-point tests. Boyle (1984) criticised Oller, finding his attitude toward discrete-point tests too negative. He claims that if discrete-point and integrative tests lie at each end of a continuum, this should suggest that one is compensating what the other one is missing, which means that one is not absolutely better than the other. Boyle further opposed Oller, who claimed that

language skills are taken apart by discrete-point tests and brought together again by the integrative tests. (p. 37)

Boyle said that Oller gave bad examples of discrete-point testing, so discrete-point testing cannot be regarded as totally "notoriously ineffective". Boyle tried to bring a reconciliation to the debate on discrete-point vs. global proficiency. He said that what is meant by discrete in discrete-point tests is not an absolute discreteness; what is actually meant is that language can reasonably be separated into its components of writing, reading, listening and speaking, and these components can be brought together with test items focusing on the nature of the language.

Besides Boyle, Zeidner (1987), who examined students' preferences for testing techniques, found that students preferred a multiple choice

examination, which is a discrete-point technique, although they believed that an essay type of examination, an integrative test, was more appropriate for reflecting one's knowledge. Zeidner's study supports Gronlund (1976) and Thorndike and Hagen (1969) who pointed out the importance of considering both "theoretical" and "practical" factors when choosing a test format.

Farhady (1983), reviewing the historical development of different types of tests which have dominated the field of second language measurement, studied the issue of integrative vs. discrete-point tests, not theoretically but statistically. He found that the actual differences between the two theoretically opposing types of tests are not as statistically distinct as has been assumed. He also claims that these tests, in fact, might be so similar that, without labeling, one could hardly distinguish them on the basis of the results.

The present study on the effects of frequent testing will be carried out by using both integrative and discrete-point tests, considering these attitudes toward discrete-point tests. Because it is clear from previous studies and discussions that discrete-point tests are statistically reliable and valid and that students prefer them, even though students do not think they are real tests of their knowledge, this study uses discrete-point tests. On the other hand, because of the criticisms of discrete-point

tests, particularly that they are not pedagogically sound and that they do not test language use adequately, a test recognized as an integrative test, was also used.

#### **2.4 Cloze Tests**

Cloze tests, which were first pioneered by Taylor (1953), are generally regarded as integrative tests. Oller (1973) describes the cloze procedure as the deletion of every nth word (fixed-ratio method) and replacing it by a blank in a paragraph, which then is presented to the learner. The learner's ability to accurately replace the deleted words or to supply appropriate alternatives is taken to be a measure of his or her linguistic competence.

Not all cloze tests are integrative tests. Cloze items could be deleted in such a way that they focus on certain discrete points of structure or morphology. Davies (1975) developed such a cloze test when he deleted only certain grammatical categories of words, for example, only function words. However, he noted that they should not be expected to be as reliable as the standard cloze tests, since these were not standard cloze items. It can also be argued that a cloze test focusing on prepositions, for example, is still an integrative test because it provides context other than the immediate sentence. Davies thinks that a focused cloze is not integrative, while an nth word cloze is, because with the nth word technique the number of words correctly

replaced or the number of contextually appropriate words supplied is a kind of overall index of the subject's ability to process the prose in context. With the deletion of certain words, such as prepositions, the learner could supply the correct answer without really understanding the meaning of the paragraph.

Oller (1973) claims that cloze tests best reflect real-life situations; therefore, he suggests that they are the only reasonable approach to testing language skills if we want to know how well the examinee can use the different elements of the language in real life communication contexts. He explains the effectiveness of integrative tests, compared with discrete-point tests, from the point of its providing meaningful contexts as follows:

The discrete-point test is a reflection of the notion from teaching theory that if you get across 50,000 structural items, you will have taught the language. The trouble with this is that 50,000 structural patterns isolated from the meaningful contexts of communication do not constitute language competence; nor does a sampling of those 50,000 discrete-points of grammar constitute an adequate test of language competence. The question of language testing is not so much whether the student knows such-and-such a pattern in a manipulative or abstract sense, but rather, whether he can use it effectively in communication. To answer the latter question, tests of integrative skills are imperative. This does not mean that discrete-point tests should never be used but that when they are used, it should be with an adequate appreciation of their practical

limitations. (p. 198)

There seems to be no research opposing the use of cloze tests as integrative tests for measuring linguistic competence. However, there are some reminders to use this technique efficiently. For example, Tarone and Yule (1989) say that if the texts which will be used as cloze paragraphs are chosen from a content area or subject matter, such as physics or chemistry, which is representative of the learner's interest, then quite specialized learner needs can be determined by means of such formats. However, when cloze tests with specific subject matters are used as diagnostic tests, they may be very hard to interpret, simply because they are so complex in terms of their content. Insufficient knowledge about the subject matter may hinder the learner from supplying the appropriate answer.

Research on the validity and the reliability of cloze tests indicate that the cloze technique is a valid and reliable means of measuring second language proficiency. Swain, Lapkin and Barik (1976) conducted research on 4th grade French-English bilingual children in order to determine the reliability and validity of cloze tests. Tests of proficiency in French were used and correlated with a cloze test in French. Proficiency tests in English were also correlated with a cloze test in English. In both cases, the correlations between the cloze scores and the other measures of proficiency used were higher

than the correlations between any of the other pairs of proficiency tests. Swain et al.'s findings were also supported by Oller (1979) who summarizes the most distinguishable features of cloze tests as follows:

Cloze tests require the utilization of discourse level constraints as well as structural constraints within sentences. Probably, it is this distinguishing characteristic which makes cloze tests so robust and which generates their surprisingly strong validity coefficients in relation to other pragmatic testing procedures. (p. 347)

However, Alderson (1979), based on research he carried out with native speakers, questioned the validity of cloze tests as a tool to measure reading comprehension and found that cloze tests relate more to test of grammar and vocabulary than to tests of reading comprehension. His study was a series of experiments carried out on the cloze procedure where the variables of text difficulty, scoring procedure and deletion frequency were systematically varied. That variation was examined for its effect on the relationship of the cloze test to measures of proficiency in English as a foreign language.

Therefore, although cloze tests are accepted as valid integrative tests, the findings of these research studies suggest that the validity of this type of test might depend on the environment in which it is applied, or the purposes for which it is used.

## 2.5 Practice Effect

When the literature for practice effect was reviewed, it was found that except for one study conducted by Schulz (1977), in which practice effect was observed only in some skills, in the other studies, researchers (Bowen, 1977; Freilich, 1989; Kirn, 1972) were not able to find any practice effect.

Freilich (1989) tried to control "practice effect" by giving a final test with a format totally different from the quizzes he gave during the treatment, in order to prevent students from obtaining higher grades just because they were familiar with the testing style. However, he found that the groups did not perform significantly differently despite treatment for test familiarity.

Another study done by Kirn (1972) who investigated whether nine weeks of biweekly practice in taking dictation (for listening comprehension purposes) and in completing cloze passages would produce a greater improvement rate in an experimental group of EFL students than in the control group from pre- to post-testing. It was found that the practice sessions did not produce any significantly greater improvement among those receiving the sessions than among those who did not.

Bowen (1977) conducted a study to investigate whether there would be a practice effect from repeated administration of a standard EFL test, using

different forms. It was found that there was no learning from the practice effect of having taken a previous form of the test.

However, a study which was carried out by Schulz (1977) at Ohio State University obtained evidence that students could enhance their performance on discrete-point items testing phonological, morphological and syntactic elements of language. One group of thirty-five students received only tests of simulated communication, and another group of forty-five students received discrete-point tests over six weeks of instruction. With respect to the discrete-point test, listening and reading comprehension were assessed through multiple-choice questions; writing and speaking were assessed through structural pattern drills. At the end of the time period, the two groups were both given a simulated communication test and a discrete-point test. The result was quite interesting. Those who had received the discrete-point treatment performed significantly better on discrete-point items testing listening, reading and writing, but not speaking. The students who had received the communication treatment did not, however, perform significantly better on communication tasks, including speaking.

The findings of the Schulz's (1977) study differ from those of the other studies. The reason that Schulz was able to observe a practice effect on some skills might be the methodology he followed. He

connected certain skills with some certain question types. For example, he tested listening and reading through multiple-choice questions, while he tested writing through structural pattern drills. His finding that practising discrete-point items (multiple choice questions) had a positive impact but that practising communication tasks did not suggests that practising certain skills with certain types of questions may help students to perform better on certain types of tests, but that the practice effect may not be universal.

## **2.6 Summary**

Researchers have conflicting views on the effects of frequent testing as reviewed in the previous sections of this chapter. Some researchers regard the diagnostic or indicative function of tests as meaningful and concrete signs of the degree of students' success in learning, and as a system for quality control. These researchers think that tests should be used together with teaching techniques. However, other researchers claim that tests are merely a burden on students' shoulders. They further believe that if tests are considered as a "must" then the amount of testing should be limited because an excessive amount of testing yields no positive results, in fact, it may impede students' performance.

Views on the practice effect seem more unified in that in the studies so far conducted, no practice

effect has been observed. The only exception was Schulz's (1977) finding that students who practised discrete-point items showed a significant increase in success while students who practised other question formats or test types showed no practice effect.

Therefore, this study aims to examine the issue of frequent testing and practice effect on BUSEL students, particularly to determine whether frequent testing through quizzes will help them increase their scores. Because there is disagreement about the effect of testing in general and frequent testing in particular, and because none of the studies reviewed are EFL studies, further exploration of the effect of frequent testing in an EFL setting seems justified.

## CHAPTER 3

### METHODOLOGY

#### 3.1 Introduction

Many researchers have focused on testing because it is seen as an important teaching tool. Some researchers have found positive educational effects arising from testing (Bloom et al., 1971; Ebel, 1980; Halpin and Halpin, 1982; Natriello and Dornbush, 1984; Peckham and Roe, 1977), but others (Bangert-Drowns et al., 1986; Freilich, 1989) have found no relation between frequent testing and obtaining better scores on a final or a post-test. Therefore, given this conflicting research, this study aimed to explore the issue once more with with EFL students and determine whether frequent testing was beneficial as discussed in section 2.3.1 or had no effect on the scores of the subjects on a post-test.

Although Freilich (1989) and Bangert-Drowns et al. (1986) hypothesised a positive relation between frequent testing and higher scores on a post-test, they found no significant effect of frequent testing on students' final scores. Bangert-Drowns et al. study showed that the number of tests given per week to the high frequency group did not have a correlation with the effect size which was a standard deviation value. Therefore, it was suggested that greater frequency of tests did not benefit the students. The study conducted by Freilich had almost the same result as that of Bangert-

Drowns et al.; that is, Freilich's three groups--quizzes for credit, quizzes for no credit and no quiz groups--received almost the same scores on the final test.

The point that distinguishes the present study from the previous studies is that the present study measured not only the effect of frequent testing, but also the practice effect and students' attitudes toward frequent testing. In both studies, Bangert-Drowns et al. and Freilich, practice effect was not taken into consideration. In addition to testing the hypothesis that there is a significant increase on students' post-test scores due to weekly discrete-point grammar and cloze quizzes, the present study also tested the following hypothesis: There is a significant relationship correlation between practising the same kind of questions on the weekly quizzes as those on the post-test and the students' gains on the sections of the post-test which contain the types of questions practised.

Although the previous two studies and the present study aimed to test the effects of frequent testing, there are some differences in their methodologies. As briefly mentioned above the Bangert-Drowns et al. study was a mega-study which actually consisted of thirty-six separate studies. In order to measure the effects of frequent testing, there were groups which received tests at different frequencies. The high frequency group received tests either once per week

or several per week. The intermediate frequency group received one test every other week. The low frequency group received one test per month. Surprisingly, results showed no significant difference in post-test scores among these three groups. However, the attitudes of the students were found to be positive toward frequent testing.

A similar conclusion was reached by Freilich (1989) who wanted to measure the effects of frequent testing in his chemistry classes. The methodology he followed differed from that of Bangert-Drowns et al.; instead of using tests at different frequencies as a variable, Freilich used one control group and two experimental groups. One of the experimental groups received tests for credit and the other for no credit and the control group received no tests at all. Freilich also found that frequent quizzing did not have any significant effect on the final scores of the students.

In the present study the effects of frequent testing were measured by comparing the scores of the control and the experimental groups on the post-test. On the other hand, the measurement of the practise effects was done by adding another dimension: the two experimental groups were given two quizzes a week, and those quizzes were made up of exercises reflecting two different testing styles. The first experimental group received structural transformation exercises as an example of discrete-

point tests and the second experimental group received cloze tests as an example of integrative tests. The practise effect was determined by comparing the performance of the subjects in the first experimental group (discrete-point test group) on the cloze portion of the test with their performance on a discrete-point subtest. The performance of the second experimental group (integrative test group) on the two subtests was also compared. It was hypothesised that each experimental group would perform better on the test items which they had practised.

At the end of the treatment, both experimental group subjects were given a questionnaire about their attitudes toward frequent testing as was done in the Bangert-Drowns et al. study.

### **3.2 Subjects**

This study was carried out at Bilkent University in Ankara, Turkey. Bilkent University is an English-medium university; therefore, the students in its various departments have to have a good command of English in order to be successful in their studies. At the beginning of each academic year both the new students and the students who failed the proficiency examination the previous year take an English proficiency exam which is prepared by the Testing Office at the Bilkent University School of English Language (BUSEL). Students who pass this examination are found to be proficient enough to enroll as

freshman students in their major departments. Students who fail the proficiency examination are given a placement test. Their placement levels are then determined as beginner, elementary, intermediate or upper-intermediate according to the scores they get on this test.

The subjects for this study were chosen from the intermediate level classes, and in order to select classes with similar levels of proficiency within this larger pool, the scores and the standard deviations of their fall term final examination, which was treated as the pre-test, were examined (see Section 3.3.1). Based on those scores, four classes were chosen within a similar acceptable range of scores and standard deviations. A questionnaire was given to the teachers of those classes to learn about their teaching styles and attitudes toward classroom testing (see Section 3.3.3), and based on the results of this questionnaire, which gave information about the teachers' tendencies toward giving quizzes or not, and their teaching strategies, three classes were chosen, one control and two experimental groups.

This study was carried out with 44 subjects: 21 females and 23 males. They were between 17 and 21 years old. The actual population of these three classes was 67. However, not all the students in these three classes were included in the treatment because the excluded students had different

backgrounds from the subjects of the study. These excluded students had been promoted or demoted from different classes which had different means and standard deviations. The subjects were not informed about the study until they were given a questionnaire to learn about their attitudes toward frequent testing (see section 3.3.4).

### **3.3 Materials**

The materials used in this study included a pre- and a post-test, which were the midterms prepared by the Testing Office at BUSEL (see Appendixes A and B), and twenty quizzes half of which were structural transformation question style quizzes and the other half of which were cloze test exercises (see Appendix C for examples). There were also two questionnaires; the first one was given to the teachers to learn about their teaching styles and attitudes toward testing (see Appendix D), and the second one was given to the students to find a correlation between their degree of success and the scores they obtained on the post-test (see Appendix E). The quizzes and the questionnaire were developed for this study.

#### **3.3.1 Pre-test/Post-test**

The midterm tests prepared by the BUSEL Testing Office were used as the pre- and the post-tests of this study. These tests were preferred by the researcher because these tests were already accepted as valid and reliable. Therefore, for practical reasons these tests were used instead of testing the

reliability and validity of newly prepared pre- and post-tests.

The final midterm of the fall semester was taken as the pre-test for the present study to determine the control and the experimental groups (see Appendix A). Likewise, the first midterm of the spring semester (see Appendix B) was taken as the post-test to measure the effects of frequent testing on the scores obtained from this test as well as the effect of practice.

These tests had five main sections: reading, vocabulary, use of English, writing, and listening. The total score of these sections was 100. There were subsections of these main sections as well. Grammar was tested in "The Use of English" section. In this section, there were four groups of questions: fill-in-the-blank questions with three options in parentheses next to the blanks (5 points), a 20 item cloze paragraph (10 points), 5 structural transformation questions (5 points), and 10 word-building questions (5 points). The total score of this section was 25.

### **3.3.2 Quizzes**

The first experimental group received discrete-point quizzes (structural transformation questions). The second experimental group received integrative quizzes (cloze paragraphs).

In BUSEL, the Curriculum Committee produces weekly syllabi for each level which set that week's

teaching schedule and focus. The content of the quizzes was designed to test what was taught in that week. Structural transformation and cloze quizzes were produced or adapted from some First Certificate Examination books (O'Connell, 1989; Jones, 1990) by the researcher. The adaptation was done to adjust the level of the quizzes through simplifying the vocabulary and sentence structure and/or eliminating the items which were above the level of the intermediate students.

Each cloze and structural transformation quiz consisted of 20 items (see Appendix C). The cloze paragraphs were selective deletion rather than every nth word. The omission of the words was done according to the grammar point tested that week. For example, in the week when the relative clauses were taught the following paragraph was given as a quiz:

The 5th of November is a day (1)  
..... children all over  
Britain light bonfires and set off  
fireworks. They are remembering Guy  
Fawkes (2)..... attempt to blow  
up the House of Parliament was  
unsuccessful in 1605. On November  
4th, Fawkes was found hiding in  
the cellars (3) ..... lie beneath  
Parliament...

Structural transformation items were written in the form of a given statement which had to be transferred to another form keeping the meaning the same. No partial credit was given for incomplete structure. For example:

- a. "Where did the robbers go?" asked the Sheriff.  
The Sheriff asked .....

- b. I will come, if it does not rain.  
Unless .....

Written instructions for the cloze and the structural transformation quizzes were provided at the beginning of each quiz. For the cloze, it was pointed out that only one word should be used per blank, and for the structural transformation, the students were reminded of the importance of providing the same meaning as the previous sentence.

### 3.3.3 Teacher Questionnaire

A questionnaire (see Appendix D) was given to the four intermediate level teachers whose classes were possible candidates for the study. It was a researcher-designed questionnaire consisting of ten questions. The questions were designed to determine whether the teachers would be willing and suitable participants in the study. There were questions asking if they were interested in participating in ELT research and asking about their attitudes toward frequent testing. There were three choices as answers: yes, sometimes/maybe, no. Teachers were supposed to tick under one of those answers. Three classes out of the four possible were chosen based on the answers given to this questionnaire by the teachers. In the selection of experimental classes special emphasis was placed upon the class teacher's willingness to co-operate. For the control group a class that usually did not receive quizzes was selected.

### **3.3.4 Student Questionnaire**

The second questionnaire (see Appendix E) was given to the students after the treatment was completed. It was given to learn whether students had found frequent testing beneficial or not. There were twelve statements with a rating scale on which the respondents were instructed to indicate their choice. The scale used in the questionnaire was a four point Likert scale: strongly disagree - disagree - agree - strongly agree. The questionnaire was given in Turkish in order to avoid any kind of misunderstanding which would have affected the results of the questionnaire and, of course, the study. Hughes (1988) justified giving a questionnaire in the students' native language on the basis of his experience in an English medium university in Turkey. He reported that the questionnaire was given in English, because it was thought that it would be inappropriate to give it in Turkish; however, it was perfectly clear that the questionnaire was largely incomprehensible to the students for whom it was intended.

### **3.4 Procedures for Data Collection**

The first step in the data collection procedure was the selection of the experimental and the control groups. That procedure was described in section 3.2.

The treatment started during the third week of the Spring term and continued for six weeks. A longer treatment period was not possible because after the

BUSEL midterm in early April, students are often assigned to new classes, so the original experimental and control groups would no longer be intact.

The first quizzes were given in the third week of the spring term and the last quizzes were given a week before the midterm. A total of twenty quizzes, ten for each group, was given to the experimental groups, while the control group received no quizzes at all. In the last week of the treatment, a questionnaire about the attitudes of the students toward frequent testing was given to the students.

During the treatment period, the experimental classes received quizzes twice a week. There was not a fixed interval between these quizzes because in order to locate the weaknesses in the students' knowledge about the point being taught, teachers tended to give the quizzes after they introduced the point or gave what they considered to be adequate instruction. The administration of the quizzes, which were designed to last 20 - 25 minutes was carried out by the teachers; however, the evaluation of the quizzes was done by the researcher.

### **3.5 Variables**

#### **3.5.1 Independent Variables**

The number and the type of quizzes were taken as the independent variables. The number of the quizzes was taken as an independent variable in order to measure the effects of the frequent testing compared to the control group which received no quizzes at all

during the treatment. The type of the quizzes, whether a discrete-point or an integrative test, provided the data in order to measure the practice effect.

### **3.5.2 Dependent Variables**

After the treatment was completed, the effects of the independent variables were measured and determined using the scores of the students on the post-test.

### **3.6 Analytical Procedures**

In order to consider the effect of frequent quizzing, practice effect and the attitude of the students' toward frequent testing, several statistical studies were done. The effect of frequent testing was determined through a series of t-test, one-way Anova and Scheffe analyses of the pre- and the post-test scores of the three groups. The t-test analysis provided the data which suggested the rejection of the experimental hypothesis on the effects of frequent quizzing. However, even with a simple inspection of the data, it was obvious that the mean score of the structural transformation group was significantly higher than the other two groups. Therefore, the gain scores between the pre- and the post-tests were calculated; and then to determine how significant the difference was a one-way Anova analysis was used. In order to do pairwise comparison of the gain scores, a post-hoc test was run, the Scheffe, which determined the effect of frequent

testing.

The practice effect on the particular sections of the test was determined by using a one-way Anova analysis. First, the gain scores of the students on these particular sections were found and then compared with each other.

The correlation between the students' attitude toward frequent testing and their post-test scores was done by using the Pearson Product Moment Correlation (PPMC).



## CHAPTER 4

### DATA ANALYSIS

#### 4.1 Introduction

The purpose of this study was to find out if a five week period of frequent testing increased the subjects' success on a post-test, and whether the practice effect and students' attitudes affected the results.

As mentioned earlier, there are several different views on the effects on frequent testing. Some researchers (Anderson and Faust, 1973; Bloom et al., 1977; Gaynor and Millham, 1976; Halpin and Halpin, 1982; Martin and Srikameswaran, 1974; Peckham and Roe, 1977) found frequent testing beneficial for the students. On the other hand, other researchers (Bangert-Drowns et al. 1986; Freilich, 1989), despite having positive views on frequent testing, ended up finding very slight or even no effect of frequent testing on the final scores of their subjects. There are yet other researchers (Davies, 1984; Noll, 1939) who expressed the view that tests were merely a burden on students' shoulders. Given these conflicting results and claims, this study aimed to explore the research goals outlined above and to determine whether it would be beneficial or not for BUSEL students to be tested frequently.

To determine the effects of frequent testing, the pre- and the post-test mean scores of each of the three groups were compared and the gain scores for

each group were calculated. The practice effect resulting from the students' practising certain types of questions during the treatment was determined through their gains on the sections in which the similar type questions they had practiced were asked. Finally, students' attitudes towards frequent testing and their post-test scores were correlated.

## 4.2 Data Analysis

### 4.2.1 Hypothesis 1

It was hypothesised that frequent testing would increase the students' degree of success on a post-test. To determine whether there was any difference caused by the frequent quizzes they had during the five week treatment period, the mean scores of pre- and post-test were calculated (Table 4.1).

**Table 4.1**  
**Means and Standard Deviations of Pre- and Post-test Scores For Each Group**

	PRE-TEST		POST-TEST	
	MEAN	SD	MEAN	SD
Group 1	69.44	7.65	58.50	7.33
Group 2	67.08	8.14	46.54	9.06
Group 3	69.00	6.79	48.31	10.36

\* Group 1: structural transformation group

\* Group 2: cloze group

\* Group 3: control group

The decrease observed from the simple inspection of the difference between the subjects' pre- and post-test scores in all three groups suggests the rejection of the experimental hypothesis. However, in order to confirm this result statistically, the

comparison of means and standard deviations of both tests of the three groups was done by using a t-test analysis (Table 4.2).

**Table 4.2**  
**T-values of the Pre- and Post-test**  
**Scores for Each Group**

Groups	T-observed	T-critical	$\alpha < 0.005$
G1	-4.38	2.89	significant
G2	-6.08	3.10	significant
G3	-6.02	3.10	significant

- \* G1: structural transformation group
- \* G2: cloze group
- \* G3: control group
- \* df: 42

Since, according to Table 4.2, there was an obvious decrease in the post-test scores of all the three groups, the results of the t-test analysis confirmed that there were significant but negative differences between the pre- and the post-test scores in each group. However, since the post-test mean values of all the groups were different from each other, the gain scores for all subjects in all three groups were calculated in order to determine how different the three groups were in terms of the magnitude of their loss from pre- to post-tests (Table 4.3).

**Table 4.3**  
**Total Gain Scores Between Pre- and Post-test**

GROUPS	GAIN SCORES
Group 1	-11.30
Group 2	-20.50
Group 3	-21.50

- \* Group 1: structural transformation group
- \* Group 2: cloze group
- \* Group 3: control group

In order to determine how significant the difference between the gain scores (see Table 4.3) of these three groups was, a one-way Anova was used (see Table 4.4).

**Table 4.4**  
**One-way Anova Analysis for the Gain Scores of Each Group**

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Btw.Grps	2	994.9419	497.4710	13.2707	.0000
Wth.Grps	41	1536.9444	37.4864		
Total	43	2531.8864			

Since the F ratio was found to be 13.2707 and F probability .0000, the results of one-way Anova indicated significantly different gain scores. To determine how significant the difference between the gain scores of the groups was a post hoc Scheffe test was used (see Table 4.5).

**Table 4.5**  
**Scheffe Test Analysis for Gain Scores**

GROUPS	OBSERVED VALUES	CRITICAL VALUE ( $\alpha < 0.01$ )
G1 & G2	4.51	3.21*
G1 & G3	4.25	3.21*
G2 & G3	0.42	3.21

G1: structural transformation group  
 G2: cloze group  
 G3: control group  
 \* : significant

As expected, the results of the Scheffe test indicated a significant difference between group 1 (structural transformation group) and group 2 (cloze group) and between group 1 and group 3 (control group). However, no significant difference was observed between group 2 and group 3.

Based on these analyses, it was found that frequent testing did not have the effect of increasing the students' scores on a post-test. There was a decrease on the mean scores of all three groups. However, the first experimental group did significantly better than the other two groups despite the decrease in the mean score of the post-test which indicated that frequent testing had significant effect on the structural transformation testing group. On the other hand, practising cloze tests did not have an effect on the cloze group's gain scores. Therefore, the null hypothesis can only be partially rejected.

#### **4.2.2 Hypothesis 2**

It was hypothesised that there is a significant

positive relationship between repeating the same kind of questions on the weekly quizzes as those on the post-test and gains on the sections in which these type of questions were asked. To determine the students' performance on these specific sections of the post-test, in which the students were asked the similar type of questions which they had practised during the treatment, the gain scores of the students in the three groups, were calculated (see Table 4.6, Table 4.7 and 4.8).

**Table 4.6**  
**Gain Scores for Structural Transformation Group on**  
**Cloze and Structural Transformation Sections**  
**on the Post-test**

Subject	Cloze	Struc. Trans.
01	0.0	3.0
02	1.5	-1.0
03	-1.5	2.0
04	-1.0	-2.0
05	-3.0	-1.0
06	-2.0	0.0
07	-2.5	3.0
08	1.0	1.0
09	-1.0	1.0
10	-0.5	-1.0
11	0.5	-2.0
12	-1.0	-2.0
13	-2.5	3.0
14	-0.5	3.0
15	2.0	-1.0
16	-2.0	4.0
17	-2.0	3.0
18	-1.0	-3.0
<b>Total</b>	<b>-15.5</b>	<b>10</b>

**Table 4.7**  
**Gain Scores for Cloze Group on Cloze and Structural Transformation Sections on the Post-test**

Subject	Cloze	Struc. Trans.
01	1.0	-2.0
02	-1.0	-3.0
03	-0.5	-5.0
04	-2.0	0.0
05	-1.0	-2.0
06	-3.5	-2.0
07	-2.5	0.0
08	-1.5	-5.0
09	1.0	-4.0
10	-2.0	-3.0
11	-1.0	0.0
12	-2.0	-3.0
13	0.0	-2.0
<b>Total</b>	<b>-15</b>	<b>-31</b>

**Table 4.8**  
**Gain Scores for Control Group on Cloze and Structural Transformation Sections on the Post-test**

Subject	Cloze	Struc. Trans.
01	-1.5	2.0
02	-1.5	-4.0
03	1.0	1.0
04	1.0	1.5
05	0.5	-3.5
06	1.0	0.0
07	-6.5	5.0
08	-3.0	-9.0
09	-1.0	-6.0
10	5.5	-3.0
11	0.5	1.0
12	-0.5	-4.0
13	0.0	-3.0
<b>Total</b>	<b>-4.5</b>	<b>-22</b>

The mean gain scores for the three groups in each section of the test is shown in Table 4.9. As can be observed from the table, the structural transformation group performed better than the cloze and the control groups.

**Table 4.9**  
**The Mean and the Standard Deviations of All**  
**Subjects in the Cloze and**  
**the Structural Transformation Sections**

	Cloze		Str. Trans.	
	Mean	SD	Mean	SD
G1	-0.86	1.42	0.56	2.25
G2	-1.15	1.31	-2.38	1.71
G3	-0.35	2.73	-1.69	3.82

G1: structural transformation group  
 G2: cloze group  
 G3: control group

In order to determine whether the cloze group had an advantage in terms of a practice effect in the cloze section of the test, a one-way Anova was used (see Table 4.10).

**Table 4.10**  
**One-way Anova Analysis for the Gain Scores on the**  
**Cloze Section of the Post-test.**

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Btw.Grps.	2	584.1540	292.0770	.835	.4411
Wth.Grps.	41	14340.2778	349.7629		
Total	43	14924.4318			

The one-way Anova gave an F ratio of .835 and an F probability of .4411 which suggests that the difference between the gain scores of these three groups was not significant. This means that none of the groups performed significantly better than the others on the cloze section of the post-test, including the second experimental group which received cloze test practice. Therefore, no practice effect was observed on the cloze section of the test for the cloze group.

Next, in order to determine whether the structural transformation group had an advantage in terms of a practice effect in the structural transformation section of the test, a one-way Anova was used to compare the gain scores of the subjects in each group in this section (see Table 4.11).

**Table 4.11**  
**One-way Anova Analysis for the Gain Scores on the Structural Transformation Section of the Post-test.**

Source	D.F.	Sum of Squares	Mean Squares	F Ratio	F Prob.
Btw.Grps.	2	8747.4553	4373.7277	6.1526	.0046
Wth.Grps.	41	29145.7265	710.8714		
Total	43	37893.1818			

Since the F ratio was found to be 6.15257 and F probability .0046, the results of the one-way Anova indicated a significant difference among the 3 groups in their gain scores on the structural transformation section of the test. This high value of F ratio meant that the difference between each group's gain scores was significant. Statistically, when a significance is found in Anova, it has to be confirmed by a post hoc test. Therefore to determine how significant the difference between the gain scores of the groups on the structural transformation section was, a post hoc Scheffe test was used (see Table 4.12).

**Table 4.12**  
**Scheffe Test Analysis for the Gain Scores on the**  
**Structural Transformation Section on the Post-test.**

Mean	Group	G2	G3	G1
-23.8462	G2			
-16.9231	G3			
7.7778	G1	*	*	

G1: structural transformation group

G2: cloze group

G3: control group

\* : significant at  $\alpha < 0.50$

For the structural transformation section of the test, there is a significant difference in the gain scores of group 1 (structural transformation group), and group 2 (cloze group) and group 3 (control group). This means that structural transformation group performed better than the cloze and the control groups in the structural transformation section of the test. However, there is no significant difference between the cloze and the control groups in terms of their performance in this section of the test.

In conclusion, the structural transformation group performed significantly better than the cloze group on the type of questions that they had practised. However, the same significant performance was not observed in the cloze group's performance in the cloze section of the test. Therefore, the null hypothesis can only be partially rejected.

#### 4.2.3 Hypothesis 3

It was hypothesised that there was a significant positive correlation between the students' post-test

scores and their attitudes toward frequent testing. Some of the questions in the questionnaire received more favorable rankings than others. The items in the questionnaire and the mean scores for these items give an idea what aspects of testing students regarded as positive (see Table 4.13).

**Table 4.13**  
**Descriptions of Items in the Questionnaire**  
**and Their Mean Scores**

Item Description	Mean Scores
#1. Frequent quizzes provide practice for the final.	3.2
#2. Quizzes make us study.	2.8
#3. Grades show how much I learn.	2.3
#4. Because of anxiety, I can't perform well on the exams.	2.4
#5. Frequent quizzes decrease test anxiety.	2.4
#6. Quizzes localise my weak points.	3.1
#7. Quizzes help me study regularly.	2.9
#8. Quizzes after each new subject.	3.3
#9. Quiz scores should be included to the final grade.	2.0
#10. Quizzes are waste of time.	1.7
#11. Similar questions on quizzes and final exam-better performance!	3.4
#12. The more practice, the higher scores.	3.4

In order to find the correlation between the post-test scores and the students' attitudes toward frequent testing, each student was given an attitude score according to their answers to the questionnaire (see Table 4.14).

**Table 4.14**  
**The Students' Post-test and Attitude Scores**

Subjects	Post-test Score (over 100)	Attitude Score (over 4)
01	55	2.8
02	55	3.0
03	53	2.4
04	65	2.5
05	62	3.0
06	73	2.5
07	50	2.9
08	62	2.8
09	50	3.1
10	74	2.8
11	54	3.1
12	66	2.6
13	60	2.9
14	58	3.3
15	57	3.3
16	49	2.8
17	54	2.4
18	56	2.5
19	51	2.7
20	40	2.5
21	52	2.6
22	52	2.6
23	40	2.6
24	60	2.5
25	62	3.0
26	46	2.3
27	49	2.8
28	29	2.9
29	40	2.4
30	43	3.1
31	41	2.5

\*  $r = 0.09$

Since the correlation coefficient was found to be 0.09, which is a non-significant value, the experimental hypothesis had to be rejected. That is, there is no correlation between the students' post-test scores and their attitudes toward frequent testing. The subjects who found frequent testing beneficial did not necessarily perform better than the subjects who did not find frequent testing beneficial. Therefore, the students' positive

attitude toward frequent testing did not have an effect of increasing their post-test scores.

#### **4.3 Interpretation and Discussion**

The data collected in a six week period in order to test the hypotheses of this study can be summarised as follows: The first aspect of the study explored the question of whether frequent testing would increase the students' rate of success on a post-test. The findings of this part of the research suggest that frequent quizzes did not have a significant positive effect on either of the experimental groups. However, it was also observed that despite the decrease of the mean scores of all the three groups, the least loss was observed in the first experimental group, the structural transformation group. The latter finding, therefore, suggests that frequent quizzes on a discrete-item type of question had some effect on the rate of success of the students.

The second aspect of the study focused on whether there is a significant relationship between practising the same type of questions on the weekly quizzes as those on the post-test and gains on the sections in which these types of questions were asked. The data obtained from this study show that the structural transformation (discrete-point) group performed significantly better than the cloze and the control groups in the section in which structural transformation questions were asked. However, the

cloze group did not show a similar performance on the cloze section of the test. That is, they did not perform significantly better than the structural transformation group or the control group although they were expected to. Therefore, the null hypothesis on the practice effect can only be partially rejected.

Finally, the third aspect of the study focused on the possibility of a significant positive correlation between the students' post-test scores and their attitude toward frequent testing. The data indicated that there was no significant correlation between students' post-test scores and their attitudes toward frequent testing. Therefore, it must be concluded that not every student who had a positive attitude toward frequent testing, performed equally well on the post-test.

The findings of this study provided data supporting those of Freilich (1979) and Bangert-Drowns et al. (1986). These researchers found that students who were tested frequently, did not perform significantly better than the untested students. Although, in the Bangert-Drowns et al.'s study, the tested group performed better than the untested group. It was not found that frequency of testing significantly influenced the academic achievement, because frequency of tests did not correlate with the success of the students. Similarly, in Freilich's study, the quiz-for-credit group did not perform

significantly better than the quiz-for-no-credit group and the control group which received no quizzes.

Therefore, the similarity observed between the present study and these two studies is that despite the expectancy of finding a positive relation between frequent tests and success of students, no such effect of frequent tests was observed.

The findings of the present study on practice effect have two aspects supporting the researchers (Bowen, 1977; Freilich, 1989; Kirn, 1972) who did not observe any practice effect as a result of their studies, and Schulz (1977), as well, who observed practice effect only on discrete-point questions. In the present study, although practice effect was observed in the discrete-point section of the test, no such effect was observed on the test as a whole. Therefore, this result supports either findings, one of which disregards practice effect and the other one which claims the presence of practice effect on discrete-point questions.

## CHAPTER 5

### CONCLUSIONS

#### 5.1 Summary of the Study

This study focused on the effects of frequent testing on students' scores on a post-test. A secondary focus was on the relationship between practising a particular question format through quizzes and the scores of the students on subsections of the post-test in which those types of questions were asked. The third focus was on the correlation between the students' post-test scores and their attitudes toward frequent testing.

The study was carried out at Bilkent University School of English Language (BUSEL). The subjects who participated in the study were from three different intermediate level classes. Two classes were treated as the experimental groups of the study and one class was treated as the control group. There was a total of 44 subjects in the study.

The treatment started with a pre-test which was actually the last midterm examination of the fall semester. The scores on this test were used to choose the three classes which were used as the control and the experimental groups. Three classes were chosen which had the closest means and standard deviations on the pre-test. The two experimental groups were given quizzes twice a week during the treatment period, which lasted six weeks, while the control group received no quizzes at all. After the treatment

was completed, but before the post-test was administered, a questionnaire was given to the subjects in the experimental groups to measure their attitudes toward frequent testing. Finally, the post-test, the first midterm of the spring term, was given.

The instruments used in the study were the BUSEL progress tests which were used as pre- and post-tests. Two different types of quizzes, discrete-point vs. integrative were given to the experimental groups. One experimental group received structural transformation questions for the discrete-point quizzes. The other experimental group received cloze paragraphs for the integrative quizzes. The third instrument was a four-point Likert scale attitude questionnaire.

There were three hypotheses related to the study. The first hypothesis stated that there was a positive relationship between frequent quizzing and the students' scores on a post-test. However, the mean scores of the students on the post-test showed that the relationship between frequent testing and the students' post-test scores were not positive but negative because there was a decrease in the mean scores on the post-test of all groups, the control and the two experimental groups. Therefore, the experimental hypothesis of the study which stated that there was a positive relationship between frequent quizzing and an increase on the students'

post-test scores had to be rejected. However, it was noticed that although all the three groups received lower scores from pre- to post-tests, the least loss was observed in the first experimental group which received structural transformation quizzes. Therefore, the gain scores of the three groups were calculated and it was found that the structural transformation group performed statistically significantly better than the other two groups in terms of their gain scores, which suggests that the null hypothesis can only be partially rejected (see 5.2 for discussion).

Similarly, when the practice effect of repeating the kind of questions similar to those on the post-test was analyzed, it was also observed that only the structural transformation group had a significant gain. Therefore, the second experimental hypothesis, that there is a significant positive relationship between repeating the same kind of questions on the weekly quizzes as those on the post-test and the gains on the sections in which these type of questions were asked could only to be partially accepted.

Further, the results of the attitude questionnaire showed that there was no positive correlation between the students' attitudes toward frequent testing and their scores on the post-test. Therefore, the null hypothesis was accepted, which meant that not necessarily every student who

perceived frequent testing as beneficial obtained a high score on the post-test.

It can be concluded from all three of these results that quizzes are not beneficial for students in terms of increasing their scores.

## **5.2 Assessment of the Study**

The results of the study about the effects of frequent testing were surprising because the discrepancies which were caused by the instruments, namely the pre- and the post-tests, were not foreseen by the researcher. These shortcomings of the instruments were realised, unfortunately, at the stage after the treatment was completed, while the data were being analyzed.

It was hypothesized that the subjects would perform better on the post-test because of the frequent quizzes they had had during the treatment period. However, the post-test mean scores of the experimental classes, which had practiced cloze and structural transformation questions, showed a decline when compared with their pre-test mean scores (see Table 5.1). This could indicate that: either the difficulty level of the post-test was higher than the previous midterms or, assuming that all three tests were equally difficult, the language level of the intermediate students in this study was lower than the two other groups of students who took the previous intermediate level tests, because their mean score was the lowest when compared to the previous

two groups.

**Table 5.1**  
**Mean Scores of Progress Tests for**  
**All Intermediate BUSEL Classes**

Test 1	Test 2	Test 3
62.59	56.10	50.86

Test 1: First midterm - Fall term  
Test 2: Final midterm - Fall term (pre-test)  
Test 3: First midterm - Spring term (post-test)

Therefore, more reliable results might have been obtained if the pre- and the post-tests were prepared by the researcher in order to insure equal difficulty levels for the tests.

The length of time, which was limited to five weeks for the administration of the quizzes, was actually a short time for this study. In a longer treatment period, there would have been the possibility of preparing the pre- and post-tests instead of using the BUSEL midterm tests, whose disadvantages were discussed at the beginning of this section. Also, the subjects were given two quizzes each week during the treatment. This might have been an overload for the students. A longer period of treatment could have enabled the subjects to evaluate the purpose of the quizzes better, so that they might have benefited more from the quizzes. Reviewing the quizzes after taking them also might have contributed to greater student success.

### **5.3 Pedagogical Implications**

The hypothesis related to the effects of frequent testing was rejected and the hypothesis on the

practice effect were only partially accepted. When the data were examined thoroughly, it was observed that the group which was given the discrete-point tests performed better than the other experimental group and the control group. The midterm examination which was used as the post-test was predominantly composed of discrete-point items and this, to a degree, explains why the discrete-point group performed better than the cloze group, because the discrete-point group found the type of questions asked on the midterm examination similar to the ones they had practised. However, the cloze group did not have that chance, because there were very few integrative type questions on the midterm examination. Therefore, as an implication of this study, it can be suggested that despite the need to reject the influence of practice partially, students should be given plenty of opportunity to practice the types of questions that will appear on a test. This might help them concentrate more on the content of the question rather than trying to understand the form of the question during the examination. It is also possible to say that practicing discrete-point items seems to be more productive than practicing integrative items.

#### **5.4 Implications for Future Research**

The initial purpose of undertaking this study was to determine if continual assessment had more positive effects on increasing students' success than

the present assessment system. Currently, students in BUSEL are assessed according to a system that does not take their midterm grades into consideration unless they are borderline students on the final examination. The score obtained on the final examination is considered to be the sole criterion for passing or failing a student. Therefore, since the final examination is crucial, any alteration of the system which allows students to demonstrate their knowledge during the semester should be considered. Insuring test familiarity and increasing the frequency of testing are two possible ways to alter the system. By giving frequent quizzes to a sample group of BUSEL students, the aim was to determine if continual assessment through quizzes would be beneficial for BUSEL students. According to the results of this study, the answer would probably be no. However, considering the factors which affected the results of the study, it would be a better idea to reconsider this issue of continual assessment once more. Under different conditions, the results of a more reliable study, which could have been done simply by writing and piloting the pre- and the post-tests by the researcher, may suggest that continual assessment would be beneficial for BUSEL students.

However, instituting continual assessment may bring new problems. If BUSEL adopted a system which assesses students continually through midterms and/or quizzes, the reliability and the validity of these

instruments of assessment may create problems. As can be seen from Table 5.1, the scores of the three tests given to the same level students at different times raises the question of the reliability and the validity of those tests. Therefore, adopting the continual assessment system would not stand as a solution by itself. This system should be established to insure the reliability and the validity of measuring students' progress by using a variety of measures which increases the chances that a students' ability has been measured accurately.



## BIBLIOGRAPHY

- Alderson, A. C. (1979). Cloze procedure and proficiency as foreign language. TESOL Quarterly 13, 219-227.
- Anderson, R. C. & Faust G. W (1973). Educational psychology. New York: Dodd, Mead & Company.
- Bangert-Drowns, R. L., Kulik, J. A., and Kulik, C. C (1986). Effects of frequent testing. (Report No. TM 860 524). Paper Presented at the Annual Meeting of the Educational Research Association (ERIC Document Reproduction Service No. ED 274 672).
- Bloom, B. S., Hastings, J. T., and Madaus, G. F. (1971). Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill.
- Bowen, J. D. (1977). Practice effect in English proficiency testing. In H. D. Brown, C. A. Yorico and R. H. Crymes (Eds.), On Tesol-77: Teaching and learning as a second language: Trends in research and practice (pp. 295-308). Washington, D.C: TESOL.
- Boyle, J. P. (1984). Being philosophical about the discrete-point versus global-proficiency debate. English Teaching Forum 22(1), 20-23.
- Bozok, S., and Hughes, A., (1984). Introduction. In Proceedings of the seminar, testing English beyond the high school. Istanbul: Bogazici University Publications.
- Carroll, B. J. (1982). Language testing - is there another way? In J. B Heaton (Ed.), Language Testing (pp. 1-10). London: Modern English Publications.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. Review of Educational Research 58, 438-481.
- Davies, A. (1984). Communicative language testing. In Proceedings of the seminar, testing English beyond the high school. Istanbul: Bogazici University Publications.
- Davies, A. (1975). Two tests of speeded reading. In R. L. Jones and B. Spolsky (Eds.), Testing language proficiency (pp. 38-104). Arlington, Virginia: Center for Applied Linguistics.

- Ebel, R. L. (1980). Evaluation of students: Implications for effective teaching. Educational Evaluation and Policy Analysis 2, 47-51.
- Farhady, H. (1983). The disjunctive fallacy between discrete-point and integrative tests. In J. W. Oller (Ed.), Issues in language testing research (pp. 311-322). Rowley: Newbury House Publishers.
- Freilich, M. B. (1989). Frequent quizzing, the final exam, and learning. Journal of Chemical Education 66, 219-223.
- Gaynor, J. & Millham, J. (1976). The Journal of Educational Psychology 68, 312-317.
- Gronlund, N. E. (1976). Measurement and evaluation in teaching. (3rd. ed.). New York: Macmillan.
- Halpin, G. & Halpin, G. (1982). The Journal of Educational Psychology 74, 32-38.
- Harris, A. C. (1975). Writing quizzes: Observations and suggestions. In Forum special issue: The art of TESOL Part 2. Washington, D.C.: USIA
- Harris, D. P. (1969). Testing English as a second language. New York: McGraw-Hill, Inc.
- Heaton, J. B. (1989). Writing English language tests. Hong Kong: Longman.
- Hughes, A. (1988). Introducing a needs based test of English language proficiency into an English medium university in Turkey. In A. Hughes (Ed.) Testing English for university study (pp. 136-146). Hong Kong: Modern English Publications.
- Hughes, A. (1989). Testing for language teachers. Cambridge: Cambridge University Press.
- Jones, L. (Ed.). (1990). Cambridge first certificate Examination practice 1. Cambridge: Cambridge University Press.
- Kirn, H. E. (1972). The effects of practice on performance on dictation and cloze tests. Unpublished MA Thesis, University of California, Los Angeles.
- Madsen, S. H. (1983). Techniques in testing. Hong Kong: Oxford University Press.
- Martin, R. R., Srikameswaran, K. J. (1974). Chemical Education 51, 485-486.

- Natriello, G. and Dornbusch, S. M. (1984). Teacher evaluative standards and student effort. New York: Longman.
- Noll, V. H. (1939). The effect of written tests upon achievement in college classes: An experiment and a summary of evidence. Journal of Educational Research 32, 345-358.
- O'Connell, S. (1989). Focus on first certificate. London: Collins.
- Oller, W. J. (1973). Discrete-point tests vs. tests of integrative skills. In J. W Oller and J. C Richards (Eds.), Focus on the learner: Pragmatic perspectives for the language teacher. Rowley, Mass.: Newsbury House.
- Oller, W. J. (1979). Language tests at school. London: Longman.
- Peckham, P. D. and Roe, M. D. (1977). The effects of frequent testing. Journal of Research and Development in Education 10, 40-50.
- Schulz, R. A. (1977). Discrete-point vs. simulated communication testing in foreign languages. Modern Language Journal 61, 94-101.
- Slavin, R. E. (1986). Educational psychology: Theory into practice. Englewood Cliffs, New Jersey: Prentice Hall.
- Swain, M., Lapkin, S., & Barik, H. (1976). The cloze test as a measure of second language proficiency for young children. Working papers in Bilingualism 11, 32-44.
- Tarone, E. & Yule, G. (1989). Focus on the language learner. Oxford: Oxford University Press.
- Taylor, W. L. (1953). Cloze procedure: A new task for measuring readability. Journalism Quarterly 33, 42-48.
- Thorndike, R. L. & Hagen, E. (1969). Measurement and evaluation in psychology and education (3rd.ed.). New York: Wiley.
- Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The students' perspective. Journal of Educational Research 80, 352 - 358.

## APPENDIX A

### TEST FORMAT OF THE PRE-TEST

The test is divided into five sections:

- \* Reading (25 points)
- \* Vocabulary (10 points)
- \* Use of English (25 points)
- \* Writing (20 points)
- \* Listening (20 points)

#### READING SECTION

There are two sections in the Reading Section:

Section A

Section B

#### Reading Section A:

This section contains 10 gap-fill sentences with multiple choice options. It samples phrases and lexis as specified in the functions and lexis lists given previously. The test format aims at assessing students' active knowledge.

e.g. How do the words behave grammatically ?

What phrases can they be used in?

What context can they be used in?

What other words can they be used with?

Some items in this section may also test students' comprehension of the stem. For example;

1. She is a very honest person. She would .....  
steal.

a. always      b. sometimes      c. often      d. never

#### Reading Section B:

One reading passage (approximately 300-350 words

in length) followed by the following question-types.

- 1. Multiple-choice comprehension questions:** These are designed to test students' overall comprehension of the text. Questions may also focus on the writer's point of view (how does the writer feel: is he for or against the idea?) and the stylistic form of the text (e.g. is the text from a newspaper, magazine, letter?). For example;
1. How does the writer feel about the changes?

a. happy      b. sad      c. surprised      d. angry

- 2. Reference Questions:** These are designed to test the students' understanding of reference systems in texts (e.g. the use of "his, the, the problem, this situation" etc.) which contribute to the cohesion of the text. Students supply a short answer. For example;

1. "This situation" in (line 5) refers to .....

- 3. Vocabulary in context:** Tests students' skill in inferring meaning of unfamiliar phrases by using overall context and textual clues. This section is generally multiple-choice in format. For example;

1. "He missed the point completely" means:
  - a. he did not agree
  - b. he did not understand
  - c. he did not want to do it
  - d. he was late for the meeting

#### **VOCABULARY SECTION**

At E4 level, students are learning certain groups

of vocabulary. These "lexical sets" can be tested in the following multiple-choice format:

**PERSONALITIES**

She always tells the truth. She is very .....

- a. talented      b. true      c. honest      d. right

**TRAVEL**

I went to Istanbul at the weekend. The ..... took seven hours.

- a. voyage      b. travel      c. journey      d. way

**USE OF ENGLISH**

In the Use of English Section there are the following exercises:

1. **3-choice options:** Tested through discrete sentences or in the context of a reading passage.

For example;

I like [ watch / watching / to watch ] TV in the evenings.

2. **Cloze test:** Students supply the missing words using only the context to help them.

3. **Transformation:** Students are asked to rewrite the sentences without changing the meaning. For example;

My sister and brother play basketball.

Both .....

**WRITING SECTION**

1. **Short-answer writing:** Completing information on a form such as a simple questionnaire.
2. **Guided writing:** Students must then transfer this information into a short paragraph.

## **LISTENING SECTION**

The material is recorded onto tapes which includes introduction, instruction and suitable pauses giving time for reading questions. There are two listenings: Listening for specific information and listening for gist may be tested both.

The following question formats are used:

- a. true / false
- b. grid-completion



## APPENDIX B

### TEST FORMAT OF THE PRE-TEST

The test is divided into five sections:

- \* Reading (25 points)
- \* Vocabulary (10 points)
- \* Use of English (25 points)
- \* Writing (20 points)
- \* Listening (20 points)

#### READING SECTION

There are two sections in the Reading Section:

Section A

Section B

#### Reading Section A:

This section contains 10 gap-fill sentences with multiple choice options. It samples phrases and lexis as specified in the functions and lexis lists given previously. The test format aims at assessing students active knowledge.

e.g. How do the words behave grammatically ?

What phrases can they be used in?

What context can they be used in?

What other words can they be used with?

Some items in this section may also test students' comprehension of the stem. For example;

1. She is a very honest person. She would .....  
steal.

a. always      b. sometimes      c. often      d. never

#### Reading Section B:

One reading passage (approximately 300-350 words)

in length) followed by the following question-types.

**1. Multiple-choice comprehension questions:** These are designed to test students' overall comprehension of the text. Questions may also focus on the writer's point of view (how does the writer feel: is he for or against the idea?) and the stylistic form of the text (e.g. is the text from a newspaper, magazine, letter?). For example;

1. How does the writer feel about the changes?

- a. happy      b. sad      c. surprised      d. angry

**2. Reference Questions:** These are designed to test the students' understanding of reference systems in texts (e.g. the use of "his, the, the problem, this situation" etc.) which contribute to the cohesion of the text. Students supply a short answer. For example;

1. "This situation" in (line 5) refers to .....

**3. Vocabulary in context:** Tests students' skill in inferring the meaning of unfamiliar phrases by using overall context and textual clues.

#### **VOCABULARY SECTION**

At L1 level, students are extending their knowledge of lexis through discussion/extension activities based around topic areas. The grammar and use of such vocabulary are to be tested in the following gap-fill format:

#### **TELEVISION**

There are seven ..... on cable TV in Turkey.

## USE OF ENGLISH

In the Use of English Section there are following exercises:

1. **3-choice options:** Tested through discrete sentences or in context of a reading passage.

For example;

I like [ watch / watching / to watch ] TV in the evenings.

2. **Cloze test:** Students supply the missing words using only the context to help them.

3. **Transformation:** Students are asked to rewrite the sentences without changing the meaning. For example;

The last time I saw her was five years ago.

I haven't .....

4. **Word Formation:** Students are asked to manipulate the word written in block capitals at the end of the sentence to fill in the blanks. For example;

I haven't made any ..... for the party.

ARRANGE

## WRITING SECTION

Writing is a separate paper. Students are required to use prompts to produce a coherent piece of text. Students should write a descriptive passage about 120-150 words.

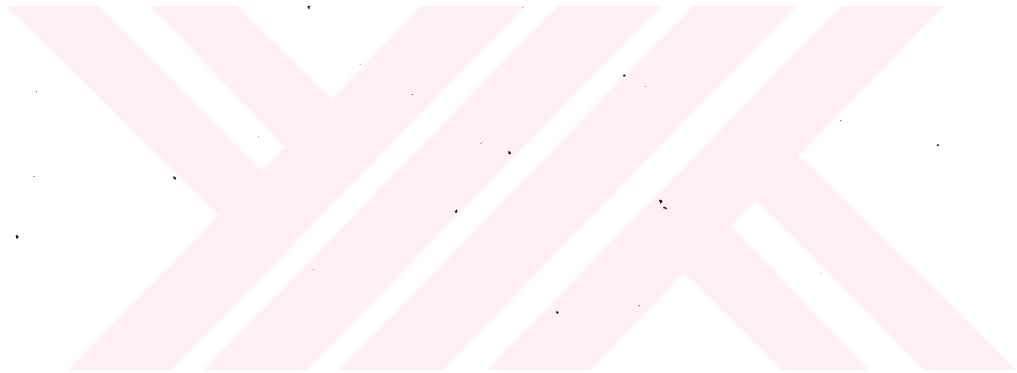
## LISTENING SECTION

The material is recorded onto tapes which includes introduction, instruction and suitable pauses giving time for reading questions. There are

two listenings: Listening for specific information and listening for gist may be tested both.

The following question formats are used:

- a. true / false
- b. grid-completion



### APPENDIX C

**Instruction :** Read this passage and complete the spaces with an appropriate word.

Romeo and Juliet are two young lovers famous throughout the world because of the play that Shakespeare wrote about their love story. The story takes (1) ..... in Verona, Italy. Romeo is a member of the Montague family (2) ..... Juliet is a member of the Capulets. For many years there has been bitter disagreement (3) ..... the two families.

When Romeo and Juliet meet, they immediately (4) ..... in love with (5) ..... other. They want to get (6) ....., but because of the feeling that exists between the two families, they realised that their wedding, as (7) ....., have to take place in secret. A priest, Friar Lawrence, performs the (8) ..... . On the day of the wedding, (9) ..... was a fight in the street and Romeo kills Tybalt, a relative of Juliet's. (10) ..... a punishment, he is ordered to go away and live in another city.

Juliet's father, not knowing Juliet's marriage to Romeo, wants her to marry someone (11) ..... . As part of the plan, Juliet, (12) ..... the help of Friar Lawrence, takes some medicine that makes it seem as (13) ..... she is dead. The plan is (14) ..... Romeo will rescue her. (15) ..... instead Romeo hears that Juliet has already died. He comes to Verona to see Juliet for the (16) .....

time, takes poison and when Juliet wakes up, finds  
him (17) ..... . She kills (18) ..... with  
a knife. The parents of (19) ..... young people  
are (20) ..... about the tragedy that they decide  
to end their quarrel.



APPENDIX C-2

**Instruction:** Rewrite the sentences with the same meaning.

1. It's three years since Jack went to Australia.

Jack .....

2. I last met Madonna two years ago.

I haven't .....

3. I haven't heard from Kevin for a long time.

The last time .....

4. It's been a long time since I saw him.

I .....

5. I last ate chocolate like this when we were in Switzerland.

I haven't .....

6. I haven't drunk coffee for years.

It's .....

7. I last saw a doctor in 1984.

I haven't .....

8. It has been a long time since I saw him.

I .....

9. We last worked together seven years ago.

We haven't .....

10. We first met three years ago.

It's .....

11. How long have you known the truth?

When .....

12. I know Tom. I first met him six months ago.

I .....

13. Jim has a beard. He grew it ten years ago.

Jim .....

14. We have been married for five years.

We .....

15. It's three months since she started learning English.

She .....

16. It's the most difficult question I've ever been asked.

I .....

17. You've never complained about it before.

It's .....

18. The last time I went out was a month ago.

I .....

19. I haven't eaten in a restaurant for three months.

The last time .....

20. It's the first time I've ever asked you for money.

I .....

**APPENDIX D**

**QUESTIONNAIRE FOR INTERMEDIATE LEVEL TEACHERS**

Name : .....

Instruction: Tick the appropriate answer for you.

QUESTIONS	YES	MAYBE	NO
1. I'm interested in innovations in ELT.			
2. I would like to participate in research done in ELT.			
3. Teaching grammar is an important aspect for me.			
4. I usually give quizzes to test my students' performance.			
5. I may allocate special time to teaching grammar while teaching a different skill.			
6. Teacher-made classroom examinations are helpful, because they help students practice for the final exam.			
7. My students are eager to learn grammar rules.			
8. Frequent quizzes increase students' test anxiety.			
9. I would like to teach all skills integrated.			
10. Intermediate level students need concrete explanations of grammar.			

**APPENDIX E**  
**QUESTIONNAIRE**

NAME : .....

INSTRUCTIONS: PLEASE ANSWER THE FOLLOWING QUESTIONS  
BY CIRCLING THE MOST APPROPRIATE ANSWER TO YOU.

1. I would like to have teacher-made quizzes frequently to have enough practice for the final examination.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

2. I think teacher-made quizzes are useful just because they make us study.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

3. I think the grades I get from quizzes show how much I learn.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

4. I usually get excited before the tests that's why I cannot perform well on the examinations.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

5. Having frequent teacher-made quizzes decreases my test anxiety.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

6. Tests help me localise my weak points, therefore I can concentrate more on those points to learn them better.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

7. I think teacher-made quizzes help me study regularly.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

8. I would like to have teacher-made quizzes after each subject so that I can determine my weak points.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

9. I think our scores on the quizzes should be included in our final grade.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

10. We should not spent our class time on quizzes, because quizzes are waste of time.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

11. If I get teacher-made quizzes with questions similar to those on the final examination, I will perform better on the final examination.

1	2	3	4
strongly disagree	disagree	agree	strongly agree

12. The more I practice, the higher scores I get.

1	2	3	4
strongly disagree	disagree	agree	strongly agree