REPUBLIC OF TURKEY
ÇUKUROVA UNIVERSITY
THE INSTITUTE OF SOCIAL SCIENCES
ENGLISH LANGUAGE TEACHING DEPARTMENT

THE EFFECT OF FORMATIVE ASSESSMENT ON LEARNERS' TEST ANXIETY AND ASSESSMENT PREFERENCES IN EFL CONTEXT

Kağan BÜYÜKKARCI

A Ph.D. DISSERTATION

ADANA, 2010

REPUBLIC OF TURKEY
ÇUKUROVA UNIVERSITY
THE INSTITUTE OF SOCIAL SCIENCE
ENGLISH LANGUAGE TEACHING DEPARTMENT

THE EFFECT OF FORMATIVE ASSESSMENT ON LEARNERS' TEST ANXIETY AND ASSESSMENT PREFERENCES IN EFL CONTEXT

Kağan BÜYÜKKARCI

Advisor :Assist. Prof. Dr. Şehnaz ŞAHİNKARAKAŞ

A Ph.D. DISSERTATION

ADANA, 2010

ÖZET

YABANCI DİL EĞİTİMİNDE BİÇİMLENDİRİCİ DEĞERLENDİRMENİN ÖĞRENCİLERİN SINAV KAYGISI VE ÖLÇME VE DEĞERLENDİRME TERCİHLERİ ÜZERİNDEKİ ETKİSİ

Kağan BÜYÜKKARCI

Doktora Tezi, İngiliz Dili Eğitimi Ana Bilim Dalı
Danışman: Yrd. Doç. Dr. Şehnaz ŞAHİNKARAKAŞ
Ekim, 2010, 123 sayfa

Sınıf içi değerlendirmenin önemi öğretmenler tarafından anlaşılmıştır; ancak, her birinin değerlendirmenin neden önemli olduğu ve ne için kullanılabileceği hakkında farklı görüşleri vardır. Bu tez, öğrenmeyi geliştirmede faydalı olan biçimlendirici değerlendirmenin, öğrencilerin sınav kaygısı ve onların değerlendirme tercihleri üzerindeki etkisini araştırmayı amaçlamıştır.

Bu çalışma kapsamında biçimlendirici değerlendirme sistemi, İngilizce Öğretmenliği bölümünde okuyan birinci sınıf öğrencilerini içeren iki farklı gruba uygulanmıştır. Bu uygulama birinci sınıf öğrencilerinin sekiz dersinden biri olan Bağlamsal Dilbilgisi Dersinde yapılmıştır. Çalışma, yapılandırıcı bir yaklaşım benimsemiş ve veri toplanması amacıyla nicel ve nitel yaklaşımlar kullanılmıştır.

Öğrencilerin sınav kaygıları ve değerlendirme tercihlerinde olan değişiklikler, uygulamanın başında ve sonunda iki farklı ölçek yardımıyla ve yüz yüze görüşme yapılarak ölçülmüştür. Çalışma bulguları, biçimlendirici değerlendirmenin öğrencilerin sınav kaygılarında olumlu değişiklik sağladığını ve genelde çoktan seçmeli testlerde yoğunlaşan öğrenci değerlendirme tercihlerinin değişmesine yol açtığını göstermiştir.

Anahtar Kelimeler: Biçimlendirici Değerlendirme, Yapılandırmacılık, Sınav Kaygısı, Değerlendirme Tercihi.

ABSTRACT

THE EFFECT OF FORMATIVE ASSESSMENT ON LEARNERS' TEST ANXIETY AND ASSESSMENT PREFERENCES IN EFL CONTEXT

Kağan BÜYÜKKARCI

Ph. D. Dissertation, English Language Teaching Department
Supervisor: Assist. Prof. Dr. Şehnaz ŞAHİNKARAKAŞ
October, 2010, 123 pages

Most would agree that the importance of assessment is recognized by teachers; however, all have different perspectives on why assessment is so important and what they can be used for. Therefore, an important part of English language teachers' expertise includes a more careful understanding of assessment. This dissertation sought the effects of formative assessment, a powerful way to improve student learning in reflective process, on students' test anxiety and their assessment preferences.

For this study, a formative assessment system was implemented into two different groups, which included same age group freshman students of English Language Teaching Department. The implementation was conducted in Contextual Grammar Course, which is one of the eight other courses of the freshman students. The study adopted a constructivist approach, and both quantitative and qualitative approaches were used for data collection.

The changes in students test anxiety and assessment preferences were elicited by two different scales and interviews both at the beginning and end of the implementation. The findings of the study showed that formative assessment had positive effects on reducing students' test anxiety, and it changed most of students' assessment preferences, which were mainly focusing on multiple choice tests.

Keywords: Formative Assessment, Constructivism, Test Anxiety, Assessment Preference.

ACKNOWLEDGEMENTS

I am most grateful to Assistant Professor Dr. Şehnaz ŞAHİNKARAKAŞ, who has supervised and helped me from the beginning of my dissertation to the end and who has given me her valuable academic advices.

I would like to express my deep gratitude to the head of English Language Teaching Department, Prof. Dr. Hatice SOFU, who has been a profound source of academic support and encouragement from the beginning of my doctorate to the end.

I am so deeply thankful to Assistant Professor Dr. Gülden and Çetin İLİN for their precious helps during my PhD and for making me feel like the other son of their family.

I would like to express my great thanks to Assistant Professor Dr. Jülide İNÖZÜ, Rana YILDIRIM, Hasan BEDİR, Oğuz KUTLU, Ahmet DOĞANAY for their support to me. Also, I would like to send my thanks to Associate Professor Dr. Erdoğan BADA who has always been not only an instructor but also a friend to me.

I am deeply indebted to Assistant Professor Cem CAN, Abdurrahman KİLİMCİ, Adnan and Münire BİÇER, Mehmet SEYİS, Fehmi Can SENDAN, Hatice ÇUBUKÇU, Neşe CABAROĞLU, Gülden TÜM, for their valuable support during my PhD.

I would like to send my thanks to Osman KELEKÇİ, Serkan DİNÇER, Özden AKYOL, Nermin ARIN and Esra ÖRSDEMİR for their friendship and support during the time I spent in Adana.(Project no: EF2008D1)

Finally, I would like to express my sincere thanks to my father, my mother, my brother, my sister and my wife, who have always backed me up whenever I needed.

TABLE OF CONTENTS

CHAPTER 3
METHODOLOGY

CHAPTER 4
RESULTS AND DISCUSSIONS

CHAPTER 5
CONCLUSIONS

LIST OF TABLES

Page

LIST OF APPENDICES

Page

CHAPTER 1

INTRODUCTION

1.0. Background to the Study

Testing or measurement is a method of evaluation of professional activities using clear criterion and frequently together with an effort at measurement either by grading on a rough scale or by assigning numerical value. On the other hand, the word assessment literally means a consideration of someone or something and a judgment about them, and it is a wider domain than testing (Brown, 2004), which can sometimes interchangeably be used with the terms testing, measurement and evaluation.

Lambert and Lines (2001) describe assessment as: a) "a fact of life for teachers, part of what teachers do; b) an organic part of teaching and learning; c) a part of the planning process." (p. 2). Erwin (1991) goes in detail in his definition of assessment as the process of collecting information on student achievement and performance, and also as the process of documenting, usually in measurable terms, knowledge, skills, attitudes and beliefs. These collected documents provide the basis for decision making regarding teaching and learning.

In spite of the variety in the way assessment is defined, it s commonly agreed that assessment is an essential part of teaching, by which teachers make a judgment about the level of skills or knowledge (Taras, 2005), to measure improvement over time, to evaluate strengths and weaknesses of the students, to rank them for selection or exclusion, or to motivate them (Wojtczak, 2002). Moreover, assessment can help individual instructors obtain useful feedback on what, how much, and how well their students are learning (Taras, 2005; Stiggins, 1992). Its systematic process provides teachers evaluating an opportunity to meaningfully reflect on how learning is best delivered, gather evidence of that, and then use that information to improve.

Regarding what components make up assessment, Marshal (2005) writes that assessment includes gathering and interpreting information about a student's performance to determine his/her mastery toward pre-determined learning objectives or

standards. Typically, results of tests, assignments, and other learning tasks provide the necessary performance data. This data can help the teacher to determine the effectiveness of instructional program at school, classroom, and individual student levels. Assessment is based on the principle that the more clearly and specifically you understand how students are learning, the more efficiently you can teach them.

When we go through the literature, we find that assessment can be classified in two main categories: The first one is *summative assessment* which is also called as assessment of learning (Stiggins, 2002; Earl, 2003). In an educational setting, summative assessments are typically used to assign students a course grade at the end of a course or project. Taras (2005) states that summative assessment is a judgment which summarizes all the evidence up to a given point. This certain point is seen as finality at the point of the judgment. This type of assessment can have various functions, such as shaping how teachers organize their courses or what schools offer their students, which do not have an effect on the learning process.

The second category is *formative assessment* that is called as assessment for learning (Stiggins, 2002; Derrich and Ecclestone, 2006). According to Black and Wiliam (1998b), assessment is referring to all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to change the teaching and learning activities in which they are engaged. Such assessment becomes formative assessment when the evidence is actually used to adapt the teaching work to meet the needs. In Threlfall's (2005) terms "formative assessment may be defined as the use of assessment judgments about capacities or competences to promote the further learning of the person who has been assessed" (p. 54).

On the basis of the views mentioned above, this study mainly focuses on students' *test anxiety* as the tests or assessment techniques utilized used in classes for grading the students are believed to create test anxiety. Most students in a class are affected from this anxiety before or during a test, and it may result in different physical consequences such as stomachache, sleeplessness, or some performance problems like students' showing a poor performance in the exams no matter how much they know.

Related research suggests that test anxiety is a mental condition in which a student experience distress before, during, or after a test or other assessment to such an extent that this anxiety causes poor performance or interferes with normal learning, and it prevents students from demonstrating their knowledge on examinations (Zuriff, 1997, Lang & Lang, 2010). For example, King et al (1991) state that there is evidence that a typical test anxious student will lose his concentration during an examination, experience difficulty in reading and understanding instructions, and also experience problems in remembering organized material; moreover, one study proofed test anxiety to be a better predictor of performance between average and high achieving students (In Putwein, 2009).

Bryan et al (1983) associate test anxiety with a strong fear of failure (In Wachelka & Katz, 1999). In other words, the students generally have the fear of getting low grades from the tests or failing a course, which is the main reason of test anxiety. However, in its traditional form, formative assessment is not generally used for grading the students; instead, it is used as an ongoing diagnostic tool by the teacher employing the results of formative assessment only to modify and adjust his or her teaching practices to reflect the needs and progress of his or her students. Thus, assessment for learning should:

1. be part of effective planning for teaching and learning so that learners and teachers should obtain and use information about progress towards learning goals; planning should include processes for feedback and engaging learners,
2. focus on how students learn; learners should become as aware of the 'how' of their learning as they are of the 'what',
3. be recognized as central to classroom practice, including demonstration, observation, feedback and questioning for diagnosis, reflection and dialogue,
4. be regarded as a key professional skill for teachers, requiring proper training and support in the diverse activities and processes that comprise assessment for learning,
5. should take account of the importance of learner motivation by emphasizing progress and achievement rather than failure and by protecting learners' autonomy, offering some choice and feedback and the chance for self-direction,

6. promote commitment to learning goals and a shared understanding of the criteria by which they are being assessed, by enabling learners to have some part in deciding goals and identifying criteria for assessing progress,
7. enable learners to receive constructive feedback about how to improve, through information and guidance, constructive feedback on weaknesses and opportunities to practice improvements,
8. develop learners' capacity for self-assessment so that they become reflective and self managing,
9. recognize the full range of achievement of all learners (Assessment Reform Group, 2002; p. 2).

In addition to the test anxiety, the study is also concerned with *assessment preferences* of the students. The study of students' preferences regarding their assessment has gained increased attention in recent years due to the growing intent of higher education institutes to adopt the service orientation (Birenbaum, 2007). The students' assessment preferences are important for understanding the factors that drive their learning process and its outcomes. For example, the previous research has shown that the assessment preference differences of the students lead to performance differences; students who preferred written assignments obtained lower marks (Watering, and et al., 2008).

According to Struyyen, Dochy, and Janssens (2002), the reviewed studies, likewise, proofed that the students' perceptions about assessment (whether summative or formative) and their approaches to learning are strongly related. For instance, Watering, Gijbel, Dochy, and Rijt explain the relation of assessment preferences with students' learning processes in their study as:

Scouller and Prosser (1994) investigated students' perceptions of a multiple choice question examination, consisting mostly of reproduction-oriented questions, to investigate the students' abilities to recall information, their general orientation towards their studies and their study strategies. The students' perceptions do not always seem to be correct: on one hand they found that some students wrongly perceived the examination to be assessing higher order thinking skills. As a consequence, these students used deep study strategies to learn for their examination. On the other hand, the

researchers concluded that students with a surface orientation may have an incorrect perception of the concept of understanding, cannot make a proper distinction between understanding and reproduction, and therefore have an incorrect perception of what is being assessed (p. 648).

Thus, the perceived characteristics of assessment seem to have a considerable impact on students' learning approaches, and vice versa. Due to the fact that students are raised as mechanical, passive and syllabus dependent learners, they barely find the chance to be involved in their own assessment processes, which gives the students very little opportunity to have different types of learning approaches.

Concerning the students' assessment preference and its relation with their learning approaches as mentioned above, one of the features that strengthen formative assessment is student involvement. If students are not involved in the assessment process, formative assessment is not practiced or implemented to its full effectiveness. Namely, students need to be involved both as assessors of their own learning and as resources to other students. This involvement helps the students be involved in assessment process and be owners of their work, and increase their motivation to learn (Black and Wiliam, 2001; Wang, 2006; Thaine, 2004).

Black and Wiliam (1998b) carried out an extensive research by reviewing of 250 journal articles and book chapters to observe whether formative assessment increases academic standards in the classroom. They concluded that efforts to reinforce formative assessment created significant learning gains as measured by comparing the average improvements in the test scores of the students involved in the innovation with the range of scores found for typical groups of students on the same tests. Effect sizes ranged between .4 and .7, and formative assessment apparently helped low-achieving students, including the ones with learning disabilities, even more than it helped other students (Black and Wiliam, 1998a).

Taking into account what the related research implies, it can be concluded that students involved in their own assessment procedure may have a lessened anxiety due to the fact that formative assessment put the emphasis on students' learning, not on their test scores. Another conclusion might be that the students' perceptions about the

assessment may have some effects on their learning approaches. As a consequence, the past experiences of the students related to assessment preferences may be changed or the students may add new perceptions of assessment by means of formative assessment usage.

1.1. Statement of the Problem

Politicians, parents, learners, and teachers have different perspectives on why assessment is so important and what they can be used for. But surely, anyone involved in language education today recognizes the importance of assessment. It has tremendous importance for teaching and learning. Fine classroom teachers use a collection of assessment tools and strategies to better understand their students' academic needs, to target their instruction, to guide next steps, and then to document their students' achievement. Assessment data informs their instruction and ensures that their teaching is open to the needs of the students. Good teachers know this and certainly connect learning with assessment (Stansbury, 2001).

The purpose of assessment is to see whether the outcomes determined in learning objectives have been reached. It can be thought in two ways: an initial state and a final state. The initial goal is the starting point, and the final goal, on the other hand, occurs when you can understand, use, or do something that you could not at the initial state. Summative assessment looks at the final product, but it does not address what occurs between the beginning and final stages of learning. Formative assessment, however, examines the progress toward the final stage (Harlen & Deakin, 2002).

Formative assessment does not get enough interest from language teachers in Turkey. One reason might be that formative assessment requires individual attention to students which might be very difficult for teachers in large classes. Another cause might be the view which accepts formative assessment as an addition to teaching, rather than an integral part of it. However, it seems that, the most important reason is the need to focus on summative assessment in classes, which is considered to lead students to success in national exams which students have to take for their further education.

In Turkey, 8 years of primary education between the ages of 6-14 is mandatory. At the end of this education period, students in general have two choices to continue their education. They can enroll in secondary schools, called Public High Schools, which do not require an exam. Or they can sit for a standard, national exam to be placed in one of the other secondary schools which provide more specific curriculum. Some of these types of schools are Anatolian High Schools, which provide more lessons in a foreign language; Science High Schools, which are focused on science education; Vocational High Schools, which focus on one type of profession.

In order to continue their education at tertiary level, every student has to take the National University Exam, a standard, high-stakes test which is organized by the Higher Education Institution of Turkey. Every year over one million people sit for this exam and less than half of them are placed at one of the universities. The competition among students results in a focus in high schools to prepare the students for universities. The students need to receive additional training by the help of private tutors and/or some other private institutions that offer preparation courses for such exams. This focus on high-stakes tests also affects the teachers' views on assessment. They want their students to succeed in these tests so they generally value assessment types that are similar to the ones in these exams. The consequence is giving priority to summative assessment, which focuses on the product, rather than the process. Additionally, the types of assessment that the students deal with are generally limited to traditional ones, such as multiple choice tests, the main type of assessment in the above-mentioned national exams.

To sum up, the students get through a long, challenging period before the national exams and as a result some students have high test anxiety levels, some have personality disorders, some become unsocial, and some even commit suicide. The researcher, however, believes that a focus not only on summative assessment but also on formative assessment might help to deal with such problems to some extent. Formative assessment applications might decrease students' test anxiety levels and might give students opportunities to experience various assessment types which might improve the quality of learning.

1.2. Aim of the Study

The aim of the current research is mainly twofold with some sub-aims. The first aim deals with the test anxiety, and the second one deals with assessment preferences of the students. According to Cassady and Gridley (2005), "students with high levels of cognitive test anxiety tend to procrastinate, worry over potential failure, utilize ineffective study strategies, and demonstrate insufficient cognitive processing skills to gain effective conceptual understanding for the content" (p. 5). Thus, the first aim of this study is to gain more information about formative assessment's effects on students' test anxiety.

Baeten, Dochy, and Struyven (2008) explain in their articles hands-on formative assessment practices have effects on students' assessment preferences. Therefore, it is also aimed to find out if implementing formative assessment causes any changes in students' assessment preferences. Shortly, this study aims to:

- *learn the effects of formative assessment on test anxiety*,
- *find out whether formative assessment leads to any changes in students' assessment preferences*.

1.3. Research Questions of the Study

Considering the purposes mentioned above, some research questions are generated for this study:

1. Does formative assessment have any effects on students' test anxiety?
2. Do students' assessment preferences change when they experience a formative mode of assessment?

1.4. Assumptions and Limitations

There are some limitations of this study. The first one is that our entire study group was the freshmen students of English Language Teaching Department, and they took the same university entrance exam and were all almost the same age group.

However, those individual variables such as age, sex, and socio-economic and cultural factors were not taken into consideration. Due to the fact that they took the same exam to be ELT department students, it was assumed that the students would perform similarly.

Another limitation is that the study focused on only four freshman classes at Cukurova University whereas there were thirteen classes during the study, which means that it would give a clearer picture if all the freshman class students could have been used as the participants in this study. The third limitation is that the study was conducted with only one teacher, the researcher himself. However, the students had eight different courses and teachers. It would be more useful if this study could have been conducted in students' other courses to see the effects of formative assessment on students' test anxiety and their assessment preferences. The last limitation is the period of the implementation. Since it is the process that is significant in this study, some more time would have been appropriate to find results that would reflect in greater accuracy.

1.5. Definitions of Terms

Constructivism: Constructivism is a theory of knowledge which argues that humans generate knowledge and meaning from their experiences. Constructivism suggests that the learner is actively involved in a joint enterprise with the teacher in the learning process. Constructivism acknowledges the learner's active role in the personal creation of knowledge, the importance of experience (both individual and social) in this knowledge creation process, and the realization that the knowledge created will vary in its degree of validity as an accurate representation of reality.

Personal Construct Theory: At the base of Kelly's theory (1955) is the image of the person-as-scientist, a view that emphasizes the human capacity for meaning making, agency, and ongoing revision of personal systems of knowing across time. Thus, individuals, like incipient scientists, are seen as creatively formulating constructs, or hypotheses about the apparent regularities of their lives, in an attempt to make them understandable, and to some extent, predictable.

Zone of Proximal Development: This term is often explained as the difference between what the learner can do without help and what s/he can do with help. Vygotsky's (1978) often quoted definition of zone of proximal development presents it as 'the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined by problem solving under adult guidance, or in collaboration with more capable peers' (p. 86).

Assessment: Assessment is an ongoing and a systematic process of looking at student success within and across courses by gathering, interpreting and using information about student learning for educational improvement (Hancock, 1994). Assessment is usually an ongoing strategy through which student learning is not only monitored--a trait shared with testing--but by which students are involved in making decisions about the degree to which their performance matches their ability. It evaluates student learning progress, and help instructors to improve the learning process (Alotaiby & Chen, 2005).

Summative Assessment (Assessment OF Learning): Assessment of Learning (AOL), also known as summative assessment, is likely to be summative and carried out periodically, like at the end of a year or a term or a unit. The teacher uses this kind of assessment to judge how well the students are performing. Conclusions will typically be reported in terms of grades, marks, or levels. According to Taras (2005), this judgement "encapsulates all the evidence up to e given point. This point is seen as the finality at the point of the judgement" (p. 468).

Formative Assessment (Assessment FOR Learning): In practice, formative assessment is a self-reflective process that intends to promote student attainment (Crooks, 2001). Formative or dynamic assessment aims at optimizing the measurement of students' intellectual abilities. They try to provide a more complete picture of child's real and maturing cognitive structures and performance and, on this basis, advance the diagnosis of learning difficulties (Allal&Ducrey, 2000).

Questioning: Effective questioning is an important aspect of the impromptu interventions teachers make once the pupils are engaged in an activity. These often include simple questions such as "Why do you think that?" or "How might you express

that?", or—in the 'devil's advocate' style— "You could argue that..." This type of questioning became part of the interactive dynamic of the classroom and provided an invaluable opportunity to extend pupils' thinking through immediate feedback on their work.

Self Assessment: Self-assessment in an educational setting involves students making judgments about their own work. Assessment decisions can be made by students on their own essays, reports, projects, presentations, performances, dissertations, and even exam scripts. Self-assessment can be extremely valuable in helping students to critique their own work, and form judgments about its strengths and weaknesses. For obvious reasons, self-assessment is more usually used as part of a formative assessment process, rather than a summative one, where it requires certification by others.

Peer Assessment: Peer assessment is assessment of students by other students, both formative reviews to provide feedback and summative grading. Peer assessment is one form of innovative, which aims to improve the quality of learning and empower learners, where traditional forms can by-pass learners' needs. It can include student involvement not only in the final judgments made of student work but also in the prior setting of criteria and the selection of evidence of achievement.

Feedback: It is described as an important part of formative assessment for both assessing pupils' current level of achievement and to show their future steps in their learning (Black and et al., 2003). Formative feedback represents information communicated to the learner that is intended to modify the learner's thinking or behavior for the purpose of improving learning.

Test Anxiety: Test anxiety is a psychological condition in which a person experiences distress before, during, or after a test or other assessment to such an extent that this anxiety causes poor performance or interferes with normal learning. While many people experience some degree of stress and anxiety before and during exams, test anxiety can actually impair learning and hurt test performance.

Assessment Preference: Assessment preference is defined as imagined choice between alternatives in assessment and the possibility of the rank ordering of these alternatives. From the studies regarding students' assessment preferences, it seems that students prefer assessment formats which reduce stress and anxiety. It is assumed, despite the fact that there are no studies that directly analyze the preferences of students and their scores on different item or assessment formats, which students will perform better on their preferred assessment formats (Watering, Gijbels, Dochy, and Rijt, 2008).

CHAPTER 2

LITERATURE REVIEW

2.0. Introduction

This chapter aims to review the relevant literature to light the way for theoretical framework of this study. The purpose is to give background information about the concepts of formative assessment within the frame of constructivism.

2.1. Constructivism

Constructivism is on the whole a theory which is based on observation and scientific study about how people learn. This theory states that people construct their own understanding and knowledge of the world through experiencing things and reflecting on those experiences. When people come across something new, they need to associate it with their previous ideas and experience (Joia, 2002). The person perhaps may change what s/he believes, or s/he may abandon the new information as irrelevant. In each case, it is the person who actively creates his own knowledge. To do this, we should ask questions, explore, and assess what we know.

The notion of constructivism has its roots in ancient times that goes back to Socrates' conversations with his followers, in which he asked directed questions that aimed his students to become conscious of the weaknesses in their thinking. The Socratic dialogue is still a significant instrument in the way constructivist teacher assess students' learning and plan new learning practices.

In our age, on the other hand, Jean Piaget and John Dewey (1944) developed some theories of child development and education which led to progress of constructivism. Piaget supposed that people learn through the construction of one logical structure after another. Lutz and Huitt (2004) state Piaget's view as:

The first aspect of Piaget's (2001) theory starts with the fact that individuals are born with reflexes that allow them to interact with the environment. These reflexes are quickly replaced by constructed mental schemes or structures that allow them to interact

with, and adapt to, the environment. This adaptation occurs in two different ways (through the processes of assimilation and accommodation) and is a critical element of modern constructivism. Adaptation is predicated on the belief that the building of knowledge is a continuous activity of self-construction; as a person interacts with the environment; knowledge is invented and manipulated into cognitive structures (p. 68).

For Dewey (1998), education depends on action. Knowledge and ideas emerge only from a situation in which learners have to draw them out of experiences that have meaning and importance to them. These situations have to occur in a social context, such as in a classroom, where students join manipulating materials and create a community of learners who construct their knowledge together. Similarly, Windschitl (2002) states that learners actively reorganize knowledge in highly individual ways, and they base rational configurations on their existing knowledge, experiences and other influences that mediate understanding.

It should be noted that constructivism itself does not suggest a particular pedagogy. Instead, it describes how learning should happen. Constructivism as an explanation of human cognition is often related to pedagogic approaches that promote active learning by doing. Constructivism thus puts the emphasis on the learner who is actively involved in the learning process, different from previous learning perspectives where the responsibility rested with the teacher to teach and where the learner played a passive, receptive role.

2.1.1. Personal Construct Theory

Initially outlined by the American psychologist George Kelly in 1955, Personal Construct Theory has been expanded to a variety of areas, including organizational development, education, business and marketing, and cognitive science. However, its predominant focus remains on the study of individuals, families, and social groups, with particular emphasis on how people organize and change their views of self and world in the counseling context.

As Donaghue (2003) states, Kelly sees a person as a scientist who tries to make sense of the universe, himself, and situations he encounters. He makes hypothesis, tests

them, and then shapes his personal constructs. These constructs are his theories or beliefs, his way of organizing and making sense of the world, and they change and are adapted with experience. Kreber and et al (2003) explain that:

The scientist's ultimate goal is seen in the prediction and control of the universe. To this end, the scientist creates working hypotheses that are then tested via experiment. If the hypothesis fails to predict/explain the actual outcome, the scientist will alter the initial hypothesis in light of the new evidence. This new hypothesis will then again be tested, and should it be verified by means of empirical assessment, the hypothesis will be considered valid for as long as it continues to accurately predict new outcomes' (p. 433).

Kelly (1955) clearly stated that each individual's psychological task is to put in order the facts of his or her own experience. Such tasks like self assessment are useful for constructing our own reality. Then each of us, like the scientist, tests the correctness of that constructed knowledge by performing those actions the constructs suggest. If the consequences of our actions are in line with what the knowledge has expected, then we have done a good job of finding the order in our personal experience. If not, then we may want to change something like our understandings or our predictions or both. The students' learning is in the same line with this theory because their experiences shape their own learning process. According to Coombs and Smith (1998), this self-organized learning has three core principles:

1. Real personal learning depends on self-assessment and reflective evaluation through the construction of internal referents;
2. the self organized learning practice depends on the ability of the learner to self-monitor and control the learning process whilst developing appropriate models of understanding; and,
3. shared meaning is negotiated conversationally from social networks. Such social networks can be understood as conversational learning environments that construct their own viability and validity, resulting in a capacity for creative and flexible thinking (In Coombs & Fletcher, 2005).

The learners, then, might be accepted as the owners of their own learning process. Being the creators of their constructs, they can reflectively consider how they understand learning as they reason on their own autobiographical lenses which frame the realities of learning (Hopper, 2000).

2.1.2. Constructivist Perspectives in Formative Assessment

In Helle's (1988) words "in society, a formative process leads to the emergence and change of social constructs; and the living reality, which forms the process, is the dynamic of interaction" (Cited in Cederman & Daase, 2003: p. 15). Thus, seen from a social perspective, society is not a static structure, but rather an emerging entity generated and constituted by an ongoing process. Just like society's not having a static structure, classroom and educational contexts cannot be considered as static. Over the last two decades, there has been a shift in the way teachers and researchers perceive student learning in education. As Nicol and Macfarlane (2006) state "instead of characterizing it as a simple acquisition process based on teacher transmission student learning is now more commonly conceptualized as a process whereby students actively construct their own knowledge and skills" (p. 199).

According to Roos and Hamilton (2004), contrary to classical test theory and its origin of the summative value of the true score which comes from behaviorist learning theories that grew in the first thirty years of the twentieth century, formative assessment, with its valued idea of feedback and development, has a different origin. It comes up from cognitive and constructivist theories of learning that emerged in the 1930s and 1940s. Shepard (2000) links formative assessment with the constructivist movement which suggests that learning is an active process, building on previous knowledge, experience, skills, and interests. Since learning is highly individualized, constructivism recognizes that teaching must be adaptive to the context, involving complex decision-making, and requiring that a teacher draw upon a collection of techniques (Giebelhaus & Bowman, 2002).

Pryor (2003) argues that "Formative assessment is better conceived of as an interactive pedagogy based on constructivist ideas about learning and integrated into wide range of learning and support activities. Instead, assessment becomes part of the

learning process so that students can play a bigger role in judging their own progress" (cited in Hangstrom, 2005: p. 26). The aim of learning is for a student to construct his or her own meaning, not just to memorize the correct answers and repeat someone else's meaning. Since education is naturally interdisciplinary, the only valuable way to measure learning is to make the assessment part of the learning process, ensuring it provides students with information on the quality of their learning.

2.1.3. Zone of Proximal Development

The concept, Zone of Proximal Development (ZPD), was developed by a Russian constructivist, Lev Vygotsky. This term is often explained as the difference between what the learner can do without help and what s/he can do with help. Vygotsky's (1978) often quoted definition of zone of proximal development presents it as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined by problem solving under adult guidance, or in collaboration with more capable peers" (p. 86).

This *actual development* (independent development) is not sufficient enough to describe the development. Rather, it points out what is achieved or developed. On the other hand, the term *proximal* (nearby) indicates that the assistance provided goes just slightly beyond the learners' current competence complementing and building on their existing abilities (Cole & Cole, 2001).

This concept, *zone of proximal development*, has tremendous implications for assessment and how we conduct assessment and evaluation. According to Puig (2003), Vygotsky criticized Western approaches to assessment for relying exclusively on estimation of the students' independent performance (zone of actual development), using fixed or summative assessment, without considering student's ability to profit from instructional interaction with more knowledgeable or capable others (zone of potential development). Vygotskians state that assessment methods must target both the level of actual development (static or summative assessment) and the level of potential development (dynamic or formative assessment) to acquire a true picture of a student's strengths and needs. Two students may have the same level of actual development, but one student may be able to solve many more problems than the other with a variety of

adult assistance. As Shepard (2005) states ‘formative assessment is a dynamic process in which supportive adults or classmates help learners move from what they already know to what they are able to do next, using their ZPD’ (66). Puig (2003) also says that Vygotsky suggests that improvements in higher order thinking skills requires social interaction within zone of proximal development, and if assessment is to have validity for learning situations that aim to promote higher order thinking skills, it should include such inter actions for sampling the student performance with ZPD. Therefore, dynamic or formative assessment seeks to assess learning within student’s ZPD.

2.2. Assessment

Clapham (2000) states that the term ‘assessment’ is used both as an umbrella term to cover all methods of testing and assessment, and as a term to distinguish alternative assessment from testing. Some applied linguists use the term ‘testing’ to apply to the construction and administration of formal or standardized tests such as the Test of English as a Foreign Language (TOEFL) and ‘assessment’ to refer to more informal methods such as those listed below under the heading ‘alternative assessment.’ For example, Valette (1994) says that tests are large-scale proficiency tests and that assessments are school-based tests. Interestingly, some testers are now using the term ‘assessment’ where they probably used the term ‘test’. There seems, definitely, to have been a shift in many language testers’ opinions so that they, probably subconsciously, may be starting to think of testing only in relation to standardized large-scale tests. They as a result use the term ‘assessment’ as the wider, more satisfactory term.

Assessment is an ongoing and a systematic process of looking at student success within and across courses by gathering, interpreting and using information about student learning for educational improvement (Hancock, 1994). The process requires that teachers think about what it is they are trying to teach, how they are teaching it, how the students learn it, what evidence shows that students are learning it, and what actions can be taken to improve student learning. The point of assessment is not to get good news, but to improve teaching and learning. Meanwhile, assessment helps to close the gap between curricular goals and student outcomes.

2.2.1. Classroom Assessment

The quality of instruction is a function of teachers' understanding of the strengths and weaknesses of their students. The depth of that understanding, in turn, depends on the quality of teachers' assessments of student achievement. Therefore, instruction requires the classroom-level assessment of student achievement. Angelo (1991) explains classroom assessment as a simple method teacher can use to collect feedback on how well their students are achieving what they are being taught.

Classroom Assessment is one method of inquiry within the framework of classroom research, a broader approach to improve teaching and learning. Research suggests that teachers spend as much as one-third to one-half of their professional time in assessment-related activities. They are continually making decisions about how to interact with the students, and those decisions are based in part on information they have gathered about their students through classroom assessment (Stiggins & Conklin, 1992). This information is usually gathered through feedback. Teacher uses feedback gained through classroom assessment to make adjustments in their teaching. Teacher also shares feedback with students to help them improve their learning strategies and study habits in order to become more independent and successful learners.

The purpose of classroom assessment is to provide teacher and students with information and insights needed to improve teaching efficiency and learning quality. Fulcher and Davidson (2007) state that performance-based elements in large-scale testing are usually limited to a small number of controlled task types. The reason for this is basically that they require considerable resources to put into practice, and are expensive. But classroom activities and assessment are almost completely performance-based, and entirely integrated. According to Ryan and Patrick (2001), this is because classroom is a social learning environment that encourages interaction, communication, achieving shared goals and providing feedback from learner to learner as well as teacher to learner.

According to Stiggins (1992), teachers use assessments to serve different purposes: to inform specific decisions, to instruct, and so on. Teachers make a lot of decisions that make instruction work when they diagnose student needs (individually

and in groups), group students for instruction, grade student performance. Each of these decisions is directly related to quality of instruction. With assessment, teachers not only make decisions, but they also instruct. That is, assessments do not just inform decisions, they are also used to teach. Teachers use assessments to inform students about their expectations and let them know what kind of skill or performance is needed to be successful. It is also helpful to use assessment for students in that when they get information about their performance through assessment, and by doing so, they can make some of their own decisions on their own learning.

2.2.2. Assessment of Learning

Assessment of Learning (AOL), also known as summative assessment, is likely to be summative and carried out periodically, like at the end of a year or a term or a unit. The teacher uses this kind of assessment to judge how well the students are performing. Conclusions of this assessment are typically reported in terms of grades, marks, or levels. According to Taras (2005), this judgement "encapsulates all the evidence up to e given point. This point is seen as the finality at the point of the judgment" (p. 468).

Harlen (2005) states that the summative uses of assessment can be grouped into internal and external to the school community. Internal uses of such assessment contain using regular grading for recordkeeping, informing decisions about courses, and communicating the results to parents and to the students themselves. Teachers' judgments, often informed by teacher-made tests or examinations, are commonly used in these ways. External uses include certification by examination bodies or for professional qualifications, selection for employment or for further or higher education, monitoring the school's performance and school accountability, often based on the results of externally created tests or examinations.

As stated before, the best way is to think of summative assessment as a means to measure, at a particular point in time, student learning relative to content standards. Although the information that is measured from this type of assessment is important, it can only help in evaluating certain aspects of the learning process. Because they are spread out and occur after instruction every few weeks, months, or once a year,

summative assessments are tools to help evaluate the effectiveness of programs, school improvement goals, alignment of curriculum, or student placement in specific programs. Black and William (1998a) state that:

All teachers have to undertake some summative assessment, for example to report to parents, and produce end-of-year reports as classes are due to move on to new teachers. However, assessing students for external purposes is clearly different from the task of assessing on-going work to monitor and improve progress (p. 143).

In other words, summative assessments happen too far down the learning path to provide information at the classroom level and to make instructional adjustments and interventions during the learning process. It takes formative assessment to accomplish this.

2.2.3. Assessment For Learning (Formative Assessment)

Learning is mostly guided by what students and teachers do in classrooms. Teachers have to control challenging situations, complicated and difficult personal, emotional, and social pressures of their students to help them learn immediately and become more successful learners. Black and William (1998a) state that teachers need to know about their students' learning progress and difficulties so that they can adapt their own work to meet pupils' needs that are generally unpredictable and vary from one student to another.

As stated before, assessment is a process to gain information about students' learning progress and their difficulties in learning, and make decisions about their students. (Black & William, 1998a; Hancock, 1994; Stiggins, 1992). This kind of assessment turns out to be formative when the evidence is used to adapt the teaching to meet students' needs.

In general terms, formative assessment, also known as *assessment for learning, on-going assessment, or dynamic assessment*, is concerned with helping pupils to improve their learning. In practice, formative assessment is a self-reflective process that intends to promote student attainment (Crooks, 2001). Cowie and Bell (1999) define it

as the bidirectional process between teacher and student to improve, recognize and respond to the learning. Similarly, Shepherd (2005) explains formative assessment as 'a dynamic process in which supportive teachers or classmates help students move from what they already know to what they are able to do next, using their zone of proximal development'(p. 66). Formative or dynamic assessment aims at optimizing the measurement of students' intellectual abilities. They try to provide a more complete picture of child's real and maturing cognitive structures and performance and, on this basis, advance the diagnosis of learning difficulties (Allal&Ducrey, 2000).

Fisher and Frey (2007) explain formative assessment and its goal as:

Formative assessments are ongoing assessments, reviews, and observations in a classroom. Teachers use formative assessment to improve instructional methods and provide student feedback throughout the teaching and learning process. For example, if a teacher observes that some students do not grasp a concept, he or she can design a review activity to reinforce the concept or use a different instructional strategy to reteach it. (At the very least, teachers should check for understanding every 15 minutes; we have colleagues who check for understanding every couple of minutes.) Likewise, students can monitor their progress by looking at their results on periodic quizzes and performance tasks. The results of formative assessments are used to modify and validate instruction (p. 4).

2.3. The Need for Formative Assessment

While many teachers are mostly paying attention on state tests, it is vital to consider that over the course of a year, teachers construct many opportunities to assess how students are learning and then use this information to make useful changes in instruction. Black and William (1998a) characterize assessment broadly to include all activities that teachers and students carry out to get information used diagnostically to alter teaching and learning. Under this definition, assessments include teacher observation, classroom discussion, and analysis of student work, including homework and tests. Assessment becomes formative when the information is used to adapt teaching and learning to meet student needs. Paul Black and Dylan William's "Inside the Black Box: Raising Standards Through Classroom Assessment" (1998a) provides

strong evidence from an extensive literature review to show that classroom "formative" assessment, properly implemented, is a powerful way to improve student learning — but summative assessments such as standardized exams can have a harmful effect.

Contrary to summative assessment, formative assessment occurs when teachers feed information back to students in ways that enable the student to learn better, or when students can engage in a similar, self- reflective process. If the main idea of assessment is to support high-quality learning, then formative assessment should be understood as the most important assessment practice.

The evidence indicates that high quality formative assessment certainly has a powerful impact on student learning. Black and William (1998a) report that the studies of formative assessment show an effect size on standardized tests of between 0.4 and 0.7, which is larger than most known educational interventions. (The effect size is the ratio of the average improvement in test scores in the improvement to the range of scores of typical groups of pupils on the same tests; Black and William recognize that standardized tests are very limited measures of learning). On the contrary, formative assessment is especially effective for students who have not done well in school, thus narrowing the gap between low and high achievers while raising overall achievement.

Ross (2005) states that one of the key appeals which formative assessment provides for language educators is the autonomy given to the learners. An advantage assumed to increase shifting the place of control to learner more directly is in the potential for the improvement of achievement motivation. Instead of playing a passive role, language learners use their own reckoning of improvement, effort, revision, and growth. Formative assessment is also thought to influence learner development through a widened area of feedback during engagement with learning tasks. Ross also says that "assessment incidents are not considered punctual summations of learning success or failure as much as an on-going formation of cumulative confidence, awareness, and self-realization learners may gain in their collaborative engagements with tasks." (p. 319).

By varying the type of assessment, as suggested above, that the teachers use over the course of the week, they can get a more accurate picture of what students know and

understand, obtaining a multiple-measure assessment window into student understanding. Additionally, by using one formative assessment daily enables educators to evaluate and assess the quality of the learning that is taking place in the classroom.

2.4. Integrating Formative Assessment in Teaching and Learning

The formal or informal assessment has some effects on teaching as well as student learning. Teachers learn about the extent to which students have learned or developed expertise, and they can tailor their teaching according to the information they gather from the students. According to Stiggins (1992), "research suggests that teachers spend as much as one-third to one-half of their valuable professional time involved in assessment related activities." (p. 211). He also states that teachers frequently make decisions about how to interact with their students and those decisions are based in part on information that they gather about their students through classroom assessment.

Informal formative assessment can occur through any teacher-student interaction. Even though teachers can not properly plan, they can prepare by making varied opportunities available for carrying out informal formative assessments such as the use of verbal interactions, questioning between teacher and students, or teachers watching and listening to students as they work through a question, problem or discussion are types of informal formative assessments that occur daily in every classroom. These assessments allow teachers to make decisions about adaptations or modifications in their instruction to promote student learning. Bachman and Palmer state that (1996) formative assessment, as an ongoing assessment, focuses on process, and it helps teachers to check the current status of their students' language ability; that is, they can know what the students know and what the students do not. It also gives chances to students to take part in adapting or re-planning the upcoming classes.

Formal formative assessments are also planned activities that are used to provide evidence about student improvement. The aims of these assessments are to inform teacher of the progress of the class, and to help the teacher check students' understanding of the elements of curriculum being taught. Black & Wiliam (1998a) suggest that for assessment to function formatively, the results have to be used to adjust teaching and learning. Whether it is classified as informal or formal formative

assessment, such assessments are essential for teachers. This is how teachers learn whether their teaching has been effective, what needs to be taught again, and when the class is ready to move on.

As on teaching, formative assessment has similar effects on students' learning. For example, this kind of assessment assists student to recognize signs from the context of study indicating what good quality work and to help them develop criteria enabling them to distinguish good from not so good performance (Boud, 2000). The evidence shows that high quality formative assessment does have a powerful impact on student learning.

Black and Wiliam (1998a) report that studies of formative assessment show an effect size on standardized tests of between 0.4 and 0.7, which is larger than most known educational interventions. (The effect size is the ratio of the average improvement in test scores in the innovation to the range of scores of typical groups of pupils on the same tests; Black and William recognize that standardized tests are very limited measures of learning.) Formative assessment is particularly effective for students who have not done well in school, thus narrowing the gap between low and high achievers while raising overall achievement.

One of the key concepts in formative assessment of teacher feedback works in two ways: to help the student improve their learning and enable the teacher to adjustments to their teaching. (Black & Wiliam, 1998a). Another concept in formative assessment is questioning. Many teachers do not plan and conduct classroom dialogue in ways that may help students learn. According to Black, Harrison and et al (2004), in this kind of dialogue, the key aspect is increasing the *wait time* which will enable students become involved in discussions and their length of their replies. Overall, teachers learn more about students' prior knowledge, and any gaps and misconceptions about that topic. Some other strategies are goal setting, self and peer assessment which will be explained in a detailed way in the following sections.

2.5. Principles in Formative Assessment

As teachers try to implement formative assessment into classroom practice, they have to decide what to try and what to develop in their context. This is because they have to make judgement about how formative assessment can be implemented within constrains of their own assessment procedures and those of their school. The Assessment Reform Group (2002) has set out 10 principles for formative assessment. According to these principles, assessment for learning should:

- be part of effective planning of teaching and learning
- focus on how students learn
- be recognized as central to classroom practice
- be regarded as a key professional skill for teachers
- be sensitive and constructive because any assessment has an emotional impact
- take account of the importance of learner motivation
- promote commitment to learning goals and a shared understanding of the criteria by which they are assessed
- enable learners to receive constructive guidance about how to improve
- develop learners' capacity for self-assessment so that they can become reflective and self-managing
- recognize the full range of achievements of all learners

Among these principles, Black and Wiliam (1998a) set out four main headings for formative assessment practice: sharing learning goals, questioning, self/peer assessment, feedback.

2.5.1. Sharing Learning Goals

Formative assessment is a systematic process to continuously gather evidence about learning. This data are used to identify a student's current level of learning and to adapt lessons to help the student reach the desired learning goal. In formative assessment, students are active participants with their teachers, sharing learning goals and understanding how their learning is progressing, what next steps they need to take, and how to take them.

Harlen (1998) lists the importance of goal setting and the strategies to be followed in target setting and goal sharing as:

*Sharing goals and expectations with pupils***:**

- It is important to find ways of communicating goals and expectations effectively with pupils.
- Using the language of aims and objectives is unlikely to be satisfactory.
- Using the teacher's criteria with students – not particularly helpful?

Two Strategies

<u>One to one with a pupil</u>*:*

- Pupil chooses 'best' pieces.
- Pupil says what she likes most and least.
- Teacher does the same, choosing key criteria, according to the substance of the work.
- Discuss differences in teachers' and pupil's perception (metacognitive reflection) generalization of knowledge and skills to other activities.

<u>Using exemplars of good quality work</u>*:*

- Discuss points that exemplify criteria of quality in concrete examples.
- Pupils gradually internalize the standards.
- Assessment genuine part of process of learning (p. 5)

Research suggests (Black, et al. 2003) that students who understand what they are being asked to learn and how they recognize success are more likely to make learning gains than those who don't. Teachers need to be clear about what they want pupils to learn from the tasks they are set. Learning opportunities are more likely to succeed if pupils have a clear, specific understanding of what they are learning and

curriculum guidelines can often help teachers to clarify the learning they want pupils to focus on.

Stiggins (2007) similarly states that one ‘key for accurate assessment is the clear, complete, and appropriate articulation of the achievement target(s) to be mastered and assessed’ (p. 15). He also says that while the learning is unfolding, students progress through the levels of proficiency leading up to each standard. For making this possible, each standard must be deconstructed into scaffolding that students must climb on their journey to that standard. These continuously unfolding classroom targets, then, become the focus of day-to-day formative assessment.

Sharing learning goals, moreover, gives the students a chance to become involved in what they are learning through discussing and deciding the criteria for success, which they can then use to recognize proof of improvements. Therefore, information about learning objectives and success criteria needs to be presented in clear, explicit language which students can understand. Quite often, messages can be expressed in language that is intelligible to the sender but meaningless to the recipient. Teachers should avoid such misunderstandings when sharing with students what they are to learn.

According to Young (2005), as well as helping pupils to be more involved in their own learning, sharing and using success criteria also provides a link into assessment of learning. If success criteria are used well, they will help pupils to identify evidence to show that they are closing the gap between where they were and where they want to be. Finding consistency in matching evidence of learning with pre-determined success criteria is also important for teachers seeking to share standards through local moderation.

2.5.2. Questioning

Questioning is another key aspect of the teaching and learning process. It is also an important element in formative assessment (Harlen, 2007). The quality of the assessment is affected by the quality of the questioning. Thus, questions should be planned and prepared so that they elicit an appropriate response from the children that

shows what they know, can do and understand. However, Black and et al. (2002) state that many teachers do not plan what to ask and say to their students. According to them, there are two main aspects in questioning: one is framing the question to be asked; and the other is timing, especially the time allowed for answering.

The quality of classroom questioning is a matter of concern, as expressed in the study of Stiggins and et al. (1989). They found that at all levels the questioning was dominated by recall questions, and while those teachers trained to teach higher-order thinking skills asked more relevant questions, their use of higher-order questions was still infrequent. An example of the overall result was that in science classrooms, % 65 of the questions was for recall, with only % 17 on inferential and deductive reasoning.

Similarly, Burns (2005) advocated questioning as a formative assessment practice which helps students to take active part in their assessment and learning. Whether verbal or written, planned questions can be used to explore student responses and elicit student reasoning. These kinds of questions provide teachers insights into student thinking that can guide their refinement of future lessons. It also helps students reflect on their own thought processes. Additionally, Black and Wiliam (1998b) identify another use of questioning to explore and develop students' prior knowledge. This kind of use requiring students to compose answers with explanations to explore their prior knowledge of new work improve learning, and this helps learners to relate the old information to the new information and to avoid superficial conclusions about the new content.

According to Forbes (2007), on the other hand, characteristics of effective questioning include matching the questions with learning target, engaging the whole class and providing enough wait time for students to respond. Moreover, all the students should be included so that the needs of both the high and low ability students are met and the bias of gender, proximity, popularity can be prevented. Besides, a study conducted by Rowe provides evidence that if wait time is about 3-5 second before giving the answer, the quality and the quantity of student responses are enhanced (Black, et al, 2003).

2.5.3. Self/Peer Assessment

Black and Wiliam (1998b) define formative assessment as referring to all actions carried out both by teachers and students in assessing themselves, which provide feedback for the future adjustment of teaching and learning activities they are engaged. Besides, according to Kelly in personal construct theory, a person is a scientist who tries to make sense of the universe, himself and situations he encounters. He makes hypothesis, tests them, and then shapes his personal constructs (Donaghue 2003). Consequently students as individual learners and as scientists of their own learning should be included in their own assessment process.

Self-assessment accordingly is a fundamental element in learning. Crooks (2001) utters that feedback on assessment cannot be effective unless students accept that their work can be improved and recognize important features of their work that they wish to develop. If students are supported to critically examine and comment on their own work, assessment can contribute powerfully to the educational development of students. Sadler (1993) similarly stated that self-assessment is essential for progress as a learner: for understanding of selves as learners, for an increasingly complex understanding of tasks and learning goals, and for strategic knowledge of how to go about improving (In Brookhart, 2001).

There are both theoretical and practical reasons for involving students in this assessment process. As stated before, the students are at the center of the process. So the more they know about -what to learn and where they are in relation with learning goals, and what further needs to be learned to reach those goals- the more they can direct their efforts usefully for learning (Harlen, 2007). There is also research showing that self-assessment can raise levels of achievement (Black and Wiliam, 1998b). In their article, Black and Wiliam explain a study's results conducted in Portuguese as:

25 Portuguese teachers of mathematics were trained in self-assessment methods on a 20-week part-time course, methods which they put into practice as the course progressed with 246 students of ages 8 and 9 and with 108 older students with ages between 10 and 14. (Fontana & Fernandes, 1994). The students of a further 20 Portuguese teachers who were taking another course in education at the time served as a

control group. Both experimental and control groups were given pre- and post-tests of mathematics achievement, and both spent the same times in class on mathematics. Both groups showed significant gains over the period, but the experimental group's mean gain was about twice that of the control group's for the 8 and 9 year-old students- a clearly significant difference (p. 10).

On the other hand, peer-assessment connects to similar principles, the most clear of which is cooperative learning (Brown 2004). It has, as Black and et al (2003) point, uniquely valuable for some reasons in students' learning; the first one is that peer-assessment has been found to improve student motivation to work carefully. Another reason is that interchange in peer discussions is in the language that the students use between themselves, which can provide common language forms, and tenable models. Third reason is that feedback from a group to a teacher can take more attention than that of an individual, and so peer-assessment helps strengthen the student voice and develops communication between teacher and the students about their learning. A further advantage is that while students are busy with their peer-assessment, the teacher can be free to observe and reflect on what is happening in the classroom. Black and et al. (2004) share a teacher's positive opinions about peer-assessment in their study:

As well as assessing and marking (through discussions and clear guidance) their own work, they also assess and mark the work of others. The students know that homework will be checked by themselves or another girl in the class at the start of the next lesson. This has led to a well-established routine and only on extremely rare occasions has failed to complete the work set. They take pride in clear and well-presented work that one of their peers may be asked to mark. Any disagreement about the answer is thoroughly and openly discussed until agreement is reached. – Alice, Waterford School (p. 14).

Harlen (2007) furthermore points that the teachers practicing student self- and peer-assessment continue that it enables students to take ownership of their learning and see themselves as partners in the teaching-learning process, raising their self-esteem. It helps them to question and filter their own concepts, and to become more self-critical.

2.5.4. Feedback

Feedback is another of the key elements of formative assessment. Vygotsky's zone of proximal development (1978) is "the distance between the actual development level by determined by independent problem solving and the level of potential development as determined under adult guidance or in collaboration with more capable peers." (p. 86; in Allal & Ducrey, 2000, p. 138). According to Ramaprasad (1983), feedback is "information about the gap between the actual level and reference level of a system parameter which is used to alter the gap in some way" (p. 4; in Sadler, 1989, p. 120). Formative assessment aims to close this distance or gap, and feedback becomes one of the basic and necessary steps.

Feedback given by the teachers is the way by which students learn about how their work match up to the expectations and how far they have met intended goals (Harlen, 2007). Black and et al (2003) similarly describe feedback as an important part of formative assessment for both assessing pupils' current level of achievement and to show their future steps in their learning. According to Wiliam and Leahy (2007), 'an assessment is formative to the extent that information from the assessment fed back with in the system and actually used to improve the performance of the system in some way' (p. 31).

The importance of feedback has been the focus of several recent research studies (Brookhart, 2007; Klecker, 2007). For example in his study, Klecker (2007) examined the impact of using feedback from multiple-choice tests on final exam scores and students Individual Development and Educational Assessment (IDEA) course ratings. He found that the mean of Class A, the class that received the weekly formative-exam feedback (M=54.45, SD=4.26), was higher than the mean of Class B, the control group (M=51.56, SD=6.48). The larger standard deviation for Class B indicated that there was greater variation in the final exam scores for the control group, and that the t-test for unequal variance should be selected to test the observed differences for statistical significance.

The feedback mentioned above might be in different ways. Butler and McMunn (2006) describe feedback as the information which can be written or oral and should

focus on helping pinpoint areas of strengths and weaknesses. It should also provide a student with information on how to improve, and it should answer the question "what do I need to know?" Moreover they state that the most effective feedback is:

- clear- easily understood and legible, if written,
- accurate- in both the student's behavior and the teacher's conclusions,
- precise- based on specific behavior, not just on generalizations,
- selective- including important observations, especially patterns of behavior noted,
- timely- given as soon as possible (p. 143).

Teachers usually use different types of feedback such as evaluative and descriptive feedback. Evaluative feedback is mostly grading. On the descriptive side, however, all of the feedback has a positive intention (Brookhart, 2008). Even criticism, if it is descriptive and not judgmental, is intended to be constructive. Descriptive feedback is telling children they are right or wrong, describing why an answer is correct, telling children what they have or have not achieved, and so on. By telling students they are right or wrong, teachers separate the correct from the incorrect; so that the students know which ways to follow in the future and which they would need to think about again. Formative feedback is even more effective when it gives details about why student's answer is correct or incorrect together with commentary on good or poor strategy. It is important to make a concerted effort to provide details such as 'this poem shows very good use of adjectives- you have used "freckly" and "polished" which is a very good description of the top side of leaves.'

Teachers commonly provide feedback to students on their work. However, if they are aware of the types of feedback they can improve its quality. According to Shute (2007), if formative feedback is to serve a corrective function, it should be able to (a) confirm whether the student's answer is correct or incorrect and (b) provide information to the learner about the correct answer. She, moreover, gives a table showing types of feedback:

Table 2.1. Feedback Types Arrayed Loosely by Complexity

Feedback type	Description
No feedback	Refers to conditions where the learner is presented a question and is required to respond, but there is no indication as to the correctness of the learner's response.
Verification	Also called *knowledge of results* (KR), or *knowledge of outcome*, it informs the learner about the correctness of her response(s), such as right/wrong or overall percentage correct.
Correct response	Also known as *knowledge of correct response* (KCR), it informs the learner of the correct answer to a specific problem with no additional information.
Try-again	Also known as *repeat-until-correct* feedback, it informs the learner about an incorrect response and allows the learner one or more attempts to answer the question.
Error-flagging	Also known as *location of mistakes* (LM), error-flagging highlights errors in a solution, without giving correct answer.
Elaborated	A general term, it refers to providing an explanation about why a specific response was correct, and it might allow the learner to review part of the instruction. It also might present the correct answer (see below for six types of elaborated feedback).
Attribute isolation	Elaborated feedback that presents information addressing central attributes of the target concept or skill being studied.
Topic-contingent	Elaborated feedback that provides the learner with information relating to the target topic currently being studied. This might entail simply re-teaching material.

Table 2.1. (Continued)

Response-contingent	Elaborated feedback that focuses on the learner's specific response. It may describe why the answer is wrong and why the correct answer is correct. This does not use formal error analysis.
Hints/cues/prompts	Elaborated feedback that guides the learner in the right direction (e.g., strategic hint on what to do next or a worked example or demonstration). It avoids explicitly presenting the correct answer.
Bugs/misconceptions	Elaborated feedback that requires error analysis and diagnosis. It provides information about the learner's specific errors or misconceptions (e.g., what is wrong and why).
Informative tutoring	The most elaborated feedback (from Narciss & Huth, 2004), this presents verification feedback, error-flagging, and strategic hints on how to proceed. The correct answer is not usually provided.

2.6. Test Anxiety

Test anxiety, which is a common problem of our time, is a feeling of fear, nervousness, and stress before a test or exam. "Anxiety, simply speaking, is a kind of troubled feeling in the mind" (Na, 2007, p. 2). As the information age continues to change, test scores have become even more important than they were in the past in evaluating candidates for admission into highly competitive educational programs (Razazadeh & Tavakoli, 2009). Therefore, many learners experience anxiety about their tests and test scores.

For the last fifty years, there have been many researches on students' test anxiety. According to Unruh and Lowe (2010), the negative relationship between test anxiety and achievement has been affirmed by multiple researchers. They also state that "test-anxious students have more difficulty in learning new material in the classroom" (p. 166). Zeidner (1998) argues in his article that "in evaluative situations associated

with desired goals, test-anxious students experience self-doubts about their ability to perform, engage in self-depreciating thoughts, and misinterpret and do not attend to relevant cues, resulting in poor test performance" (In Lowe et al, 2008. p. 217).

The reaction of the highly test anxious individuals may include a range of unpleasant and debilitating anxiety symptoms. Zuriff (1997) identifies three components:

The first, subjective distress includes fear or apprehension, sometimes bordering on terror, dizziness, tension, stomach discomfort and a fear of fainting. The second, physical symptoms, includes trembling, sweating, changes in heart rate and breathing, clammy hands, voice tremor and muscular tension. These are accompanied by distracting, intrusive negative thoughts such as 'I cannot answer this question', 'I am going to fail this course' or 'my life is a failure'. The third refers to cognitive effects; concentration becomes difficult, examination questions seem unrecognizable and the person experiences their mind 'going blank' (In Putwain, 2009, p. 57).

In order to lessen these kinds of effects of test anxiety, there have been some different ways proved to be effective in a lot of studies. For example, Wachelka and Katz (1999) found that "a cognitive-behavioral treatment using relaxation training, guided imagery, self-instructional training, and study skills training reduced test anxiety." (p. 197). In another study, Cassady and Gridley (2005) found that "allowing students to freely access practice tests and receive immediate corrective feedback provides personal control over test preparation" (p.24), which may help reducing test anxiety.

2.7. Assessment Preference

It is commonly known that assessment has a crucial role in students' learning process. The way students prepare themselves for an assessment largely depends on how they see the assessment, and this can affect their learning positively or negatively (Watering, Gijbels, Dochy, and Rijt, 2008). The development and implementation of teaching practices that will promote students to obtain and apply their knowledge efficiently, think critically, analyze, synthesize, and make inferences are the important challenges for today's higher education (Gijbels, Segers, & Struyf, 1998). It is said that

students should adopt more deep approaches to learning in order to achieve these challenges. In general, it is claimed that new learning environments have the potential to improve these educational outcomes for students in higher education by making the students' learning the core subject and defining instruction as enhancing learning (Dart 1997; Lea et al, 2003).

Clarke, Heaney and Gatfield (2005) assume that students in all countries prefer multiple-choice testing, and they link this to a widespread perception of ease, a lesser risk of failure on the basis of linguistic expression, and the statistically favorable potential of blind guessing (In Bartram & Bailey, 2010). However, Birenbaum (2007) argue that students preferring problem-solving tasks tend to perform better than students preferring simple, quick, and easy problems. Besides, since new assessment methods are largely available in the last decades, students' assessment preferences are not restricted to their experiences of multiple-choice examinations and essay modes of evaluation anymore (Struyven, Dochy, and Jassens, 2005).

It can be said that the students' preferences through instruction and assessment can also reflect their perceptions of learning environment, and their approaches to learning which can affect their success. Birenbaum (1997) states in her study that if the students are provided with "the assessment type they prefer from among those types of assessment considered appropriate for a given purpose, the perceived validity of the assessment will improve, thus motivating them to perform at their best" (p. 81). To sum up, the assessment preferences of the students, as explained above, have some effects not only their learning preferences but also their achievements.

CHAPTER 3

METHODOLOGY

3.0 Introduction

This chapter firstly examines the mixed methods (qualitative and quantitative) used in this study to explain its main questions about the effects of formative assessment on students test anxiety, assessment preferences. Using both qualitative and quantitative methods for collecting data for this study enabled the researcher to explore the perceptions of the students about formative assessment practices, and to help in clarifying the nature of their insights about the formative assessment practices.

Secondly, the participants of this study are discussed in terms of their sex, age, and their backgrounds, and the rationale for choosing these participants, and this course. Next, the reasoning for the data collection tools is explained, and those data collection tools and their validation studies are illustrated separately. Furthermore, data collection procedure is explained at the end of this chapter. This procedure clarifies the steps of pilot study, which helped the researcher to finalize both the application process and the data collection tools' latest versions. Lastly, the phases of the main study are examined regarding what was done for formative assessment applications, and how the data was collected.

3.1. Design of the Study

The study is based on a constructivist theory which argues that people produce knowledge and meaning from their experiences (Piaget, 2001). The theory of constructivism is vital for formative assessment because it conceptualizes learning as an active process, which builds on students' own understanding. Meanwhile, the central principle of this study is that students should be involved in their own assessment process, and by means of this participation, there may be some changes in students' test anxiety. Moreover, they may increase their understandings and add new types of assessments to their preferences.

To provide these aims, the mixed methods were included using quantitative and qualitative strategies aiming to collect and analyze both forms of data in a single study (Cresswell, 2003). Being aware of that all methods have their own limitations, researcher felt that triangulating data sources- through using quantitative and qualitative methods- would lessen the danger of using only one source with its biases. During this study, concurrent procedures, in which the researcher integrated quantitative and qualitative data in order to provide a clear analysis of the research problems, helped the researcher collect both forms of data at the same time and then to integrate the information for the explanation of the overall results. Also, in this design, the researcher united one form of data within another to analyze different questions of the study.

3.1.1. Quantitative Data

In quantitative research, the aim is to determine the connection between one thing (an independent variable) and another (a dependent or outcome variable) in a population (Hopkins, 2000). Quantitative research includes two types of strategies of inquiry: experimental and surveys. Surveys include longitudinal studies using questionnaires or structured interviews for data collection. Experimental studies, on the other hand, are known as *longitudinal* or *repeated-measures* studies for obvious reasons. They are also referred to as *interventions* because you do more than just observe the subjects.

In this study, the effects of formative assessment with this study are sought on experimental group students' test anxiety and assessment preferences. Therefore, those students are not only observed, but they are also subjected to a formative assessment treatment for fourteen weeks. In addition, the researcher collected data on predetermined instruments such as test anxiety inventory, and assessment preference scale, which yielded statistical data. The selection and validation studies of the quantitative data collection tools are explained in detail in the following sections.

3.1.2. Qualitative Data

"The expression qualitative methodology refers in the broadest sense to research which produces descriptive data, people's written or spoken words and observable behavior" (Taylor & Bogdan, 1998, p. 7). Typically, when a longitudinal view is used in

educational research, the study either uses survey data to examine students' beliefs before and after a particular treatment ,or the study uses a qualitative approach following a small number of students through their student preparation program, focusing on their unfolding conceptions of teaching or learning (D'Souza, 2009). Due to the fact that this study, similarly, aimed to unfold students' test anxieties and their assessment preferences, field notes, semi-structured interviews (pre and post), and teacher observations were used both at the beginning and at the end of the treatment.

Also the researcher completed a more holistic and qualitative analysis focusing on what different formative assessment activities the teacher provided, how students perceived those assessment practices, whether teacher provided written and verbal feedback, whether students responded to this feedback, and whether students provided feedback to themselves and their classmates.

3.2. Participants

The participants of this study were the freshmen students of Cukurova University English Language Teaching Department. The participants consist of thirty eight experimental group students and forty eight control group students, and eighty six students in total. The department has thirteen freshmen classes. Both experimental and control groups have two classes, meanwhile, four classes in total took part in this study. The experimental group consists of nine boys and twenty nine girls; on the other hand, the control group has eight boys and forty girls. Their ages vary from eighteen to twenty.

Two classes for experimental group and two classes for control group (four in total) were randomly chosen out of thirteen classes because with random selection or random sampling, each student has an equal chance to be selected from the population, guaranteeing that the sample will be representative of the population (Creswell, 2003). All of the students have been learning English since 4^{th} grade of primary school, which means for at least 8 years. Moreover, all of them took the university entrance exam last year, 2009, to be a student at Cukurova University. Besides, both experimental and control group students had one year preparatory class in their first year of the university.

The formative assessment treatment was applied in students' "Contextual Grammar" course because it seems that the students take it as the *hardest* one among other courses. This is due to the fact that before coming to university all the students had to take university entrance exam that was a multiple choice high stake test. The students spent, at least, their last two years by only working on multiple choice tests, which means they did not have to produce any piece of work, like speaking or writing. However, this course deeply embedded in language usage, and rhetorical grammar, which require using and producing language. Therefore, they were not feeling so secure about this course, and had fears about the assessment of the course. This issue was clear in their test anxiety inventory and interview results which are discussed in the data analysis chapter.

3.3. Data Collection Tools

Data of this study were collected from different sources for a better understanding of the changes in students' test anxiety levels and their assessment preferences. For this purpose, the first tool for data triangulation was the *Test Anxiety Inventory* that was used to reveal how students were seeing the tests in their lives, and what kind of effects those tests had on their body and thoughts at the beginning of the study, and to learn what kind of changes occurred during and after the study.

For a deeper understanding, another tool, *Assessment Preference Scale,* was applied before and after the formative assessment implementation. It helped the researcher to find out what kind of assessment preferences the students had, and why they had those preferences; then at the end of this process, this scale showed the changes, if any, in the students' assessment preferences.

Additionally, as a data collection tool, semi-structured interviews were conducted with randomly selected seven students of the experimental group before and after the formative assessment implementation. The first interview showed why students had fears (if any) about tests and classroom assessment, and it also showed what kind of assessments they preferred before this study process. The second interview, on the other hand, helped the researcher to find out what differences occurred in the ideas of the students about their test anxiety and their classroom assessment preferences. Lastly,

field notes and teacher observations helped the researcher to get more detailed information about formative assessment practices in the classroom. The following sections explain these data collection tools in detail.

3.3.1. Test Anxiety Inventory

Test anxiety is a psychological situation in which a person experiences distress before, during, or after a test or other assessment to such an extent that this anxiety causes poor performance or impede student learning (Shahi, 2009). A U.S. Department of Education-funded study was conducted by the Institute of HeartMath and Claremont Graduate University with 980 10^{th}-grade students. The study found a strong negative relationship between test anxiety and test performance: Students with high levels of test anxiety scored, on average, 15 points lower on standardized tests in both mathematics and English-language arts than students with low test anxiety (Institute of HeartMath, 2007). Therefore, it was necessary to see students' test anxiety before and after formative assessment implementation of the current study.

For this aim, the Test Anxiety Inventory (TAI) is used as the first instrument of the study. It is a multi-dimensional measure used to assess the test anxiety of freshmen students of the study. TAI is a self report (see table 3.1 for examples) and consists of 50 items on a 2-point scale where *1 = false* and *2 = true*. It has four main sources of test anxiety- *concerns about: a) how others will view you if you do poorly* (8 items)*, b) your own self image* (7 items)*, c) your future security* (6 items)*, d) not being prepared for a test* (6 items). It also includes three main expressions of test anxiety*: a) bodily reactions* (7 items)*, b) thought disruptions* (10 items)*, c) general test-taking anxiety* (6 items).

The highest anxiety score that can be obtained for each category is calculated by multiplying all the items of that category with 2, which is the maximum anxiety score given for *true*. For example, the first category (others' views) has eight items and it can have sixteen as the maximum score for its anxiety level.

Table 3.1. Examples of Items about Main Sources and Expressions of TAI for Students

ITEM NO	**ITEM**	**MAIN SOURCES and EXPRESSIONS**
3	People (family, friends, etc.) are counting on me to do well.	Concerns about how others will view you if you do poorly
10	Event though I don't always think about it, I am concerned about how others will view me if I do poorly.	Concerns about how others will view you if you do poorly
2	Getting a good score on one test does not seem to increase my confidence on other tests.	Concerns about your own self
9	Before or during an important exam, I find myself thinking about how much brighter some of the other test-takers are.	Concerns about your own self
1	I wish there were some way to succeed without taking tests.	Concerns about your future security
8	People who do well on tests generally end up in better positions in life.	Concerns about your future security
6	I have always dreaded courses in which the teacher has the habit of giving pop quizzes.	Concerns about not being prepared for a test
11	Worrying about how well I will do interferes with my preparation and performance on tests.	Concerns about not being prepared for a test
5	I do not enjoy eating before or after an important test.	Bodily reactions
12	Having to face an important test disturbs my sleep.	Bodily reactions
4	During a test, I sometimes find myself having trains of thought that have nothing to do with the test.	Thought disruptions
13	I cannot stand to have people walking around watching me while I take a test.	Thought disruptions
7	It seems to me that test sessions should not be made the formal, tense situations they are.	General test-taking anxiety
14	If exams could be done away with, I think I would actually learn more from my courses.	General test-taking anxiety

As it is clear from the table, this scale helped the researcher to reveal the possible effects of tests such as how others will view them, their self confidences, future security, and their bodily and thought disruptions. Although the main aim of the treatment in this study was not to proof whether formative assessment would lessen all

those student test anxieties, TAI was also important in terms of showing whether formative assessment would be useful in reducing some of those anxieties of the students.

TAI was adapted according to the data collected after the pilot study. Each category was analyzed by means of factor analysis separately in order to get homogenous categories, and after several analysis of each category, some of the items in the categories had to be removed from the inventory to get a higher validity. The process of removing items in each category is as follows:

- *First category (Others' Views)*: This category included 8 items at the beginning. When those items were subjected to factor analysis, they gathered under three factors (dimensions), and the first factor explained only % 30.075 of the total variance. To get the items under one factor, the ones whose factor loads were less than .50 (items 3, 10, 25, 32) were extracted. Then the rest four the items loaded on one factor, and explained 52.025 of the variance. Their factor loads were between .60 and .76. The reliability of this category was .69 after factor analysis.

- *Second Category (Self-Image):* The second category included 7 items before the analysis. When the items were subjected to factor analysis, they gathered under four factors, and the first factor explained % 19.153 of the total variance. To gather the items under one factor, the ones whose factor loads were less than .50 (items 9, 16, 24, 38) were extracted. Then all the rest of the items were under only one factor and explained 43.161 of the variance. The rest of the items' factor loads were between .60 and .75. The reliability of this category was .53 after factor analysis.

- *Third Category (Future Security)*: The third category included 6 items at the beginning. When those items were subjected to factor analysis, they gathered under three factors, and the first factor explained % 24.661 of the total variance. To get the items under one factor, the ones whose factor loads were less than .50 (items 1, 30) were extracted. Then all the rest of the items were under only one factor and explained 36.004 of the total variance. The rest of the items' factor loads were between .55 and .65. The reliability of this category was .56 after factor analysis.

- *Fourth Category (Being Prepared For A Test)*: The fourth category included 6 items at the beginning. When those items were subjected to factor analysis, they gathered under two factors, and the first factor explained % 31.047 of the total variance. To get the items under one factor, the ones whose factor loads were less than .50 (items 6, 11, 26) were extracted. Then all the rest of the items were under only one factor and explained 52.447 of the variance. The rest of the items' factor loads were between .68 and .73. The reliability of this category was .55 after factor analysis.

- *Fifth Category (Bodily Reactions)*: The fifth category included 7 items at the beginning. When those items were subjected to factor analysis, they gathered under three factors, and the first factor explained % 27.707 of the total variance. To get the items under one factor, the ones whose factor loads were less than .50 (items 19, 27, 39) were extracted. Then all the rest of the items were under only one factor and explained 44.213 of the variance. The rest of the items' factor loads were between .69 and .72. The reliability of this category was .58 after factor analysis.

- *Sixth Category (Thought Disruptions)*: The sixth category included 10 items at the beginning. When those items were subjected to factor analysis, they gathered under four factors, and the first factor explained % 25.490 of the total variance. To get the items under one factor, the ones whose factor loads were less than .50 (items 13, 21, 35, 48) were extracted. Then all the rest of the items were under only one factor and explained 36.090 of the variance. The rest of the items' factor loads were between .55 and .60. The reliability of this category was .63 after factor analysis.

- *Seventh Category (General Test-Taking Anxiety)*: The seventh category included 6 items at the beginning. When those items were subjected to factor analysis, they gathered under three factors, and the first factor explained % 26.561 of the total variance. To get the items under one factor, the ones whose factor loads were less than .50 (items 29, 44, 45) were extracted. Then all the rest of the items were under only one factor and explained 50.735 of the variance. The rest of the items' factor loads were between .70 and .71. The reliability of this category was .51 after factor analysis.

After removing those items above from TAI, the rest of the items (27 items in total) were used for the main study, and the Cronbach's alpha of the adapted inventory was found to be .80, which shows the total reliability of the seven categories.

3.3.2. Assessment Preference Scale

The assessment methods in use in higher education have expanded noticeably in recent years. More recently, portfolios, self and peer assessment, and other new methods have been introduced in higher educational contexts. Students' understanding about these recent formats of assessment and the more common multiple-choice and essay examinations constitute an important part of this study.

The studies on students' assessment preferences have also gained an increased attention in recent years because it is important to understand the factors which drive the learning process and its outcomes (Birenbaum, 2007). Previous research has shown that differences in students' learning and testing lead to performance differences, and students' assessment preferences reflect their perception of learning environment, their approaches to learning, and consequently, affect their achievement (Biggs, 2003, Struyven & et al, 2005).

In this study, assessment preferences of the students' are considered as an important part of the study. The effects of formative assessment implementation on students' assessment preferences are expected to change from a traditional assessment type, one of the most common type of assessing student skills and achievement, to a formative type of assessment which leads to a real production of student achievement.

For the aim of discovering students' assessment preferences, Assessment Preference Scale has been developed and used as a data collection tool both at the beginning and end of the study. The steps of developing this scale are as follows:

1. The relevant literature was examined, and 16 items related to assessment preferences were composed. Three assistant professors in English Language Teaching department and two experts in measurement and evaluation were asked to examine the comprehensibility and sufficiency of the items. It

consisted of a 5-point scale where *1 = never*, *2 = rarely, 3 = sometimes, 4 = often,* and *5 = always*.

2. The scale was applied to Cukurova University English Language Teaching Department freshman students (n = 107) aging from 18-20 in 2008-2009 educational year.
3. Factor analyze was conducted in order to determine reliability of the scale. After these steps, this data collection tool was determined to be a 16 item scale.

The analysis showed that the scale had a high reliability (Cronbach's alpha= .84). This scale has two main categories: *a) Traditional Assessment Preferences* including the sub-categories as*:*

- *selected response tasks (multiple-choice, matching),*
- *production tasks (open-ended questions, explanatory questions),*
- *limited-production tasks (rewrite, clause test, combining sentences)*,

b) *Formative Assessment Preferences* including the sub-categories as:

- *self/peer assessment*,
- *feedback preferences (verbal, written),*
- *ongoing assessment preferences (weekly quizzes, classroom activities, take-home task*.

The highest assessment preference score that can be obtained for each category is calculated by multiplying all the items of that category with 5, which is the maximum anxiety score given for *always*. For example, the first category (traditional assessment preference) has seven items and it can have thirty five as the maximum score for its assessment preference level.

3.3.3. Interviews

An interview is a talk between the interviewer and the interviewee where questions are asked by the interviewer to obtain information from the interviewee. As

Taylor and Bogdan (1998) state that social scientists mostly rely on spoken accounts to learn about social life. In social research interviews, the purpose is to elicit information from respondent or interviewee about their attitudes, norms, beliefs, and values (Bryman, 2008). Shortly, the aim of an interview is to find out what is in people's mind.

According to Bryman, there are four types of interview: structured, semi-structured, informal, and retrospective. Structured and semi-structured ones are verbal questionnaires. They consist of series of questions designed to obtain information from respondents. Informal interviews are much less formal than the first two interview types. They intend to find out what people think and how the views of one individual compare with those of another. Retrospective interviews may be structured, semi-structured, or informal. It aims to get a respondent to recall and then reconstruct from memory something happened in the past.

In this study a semi-structured interview has been conducted both at the beginning of the study and at the end of formative assessment implementation. The reason for using semi-structured interview is that it is more flexible than standardized methods such as the structured interview or survey. Although the researcher in this study had some established general topics for examination, this method allowed for the exploration of rising ideas rather than just relying only on thoughts and questions defined in advance of the interview.

A standardized interview schedule with set questions asked of all respondents was used by the researcher. The questions were asked in a similar order and format to make a form of comparison between answers possible. On the other hand, there was also scope for pursuing and investigating for new and related information, through additional questions. Also, impromptu questions were formulated in order to follow up some unexpected emerging ideas during the interview. The followings are the two examples of the interview questions:

1- When you were in high school, how did your teachers assess you in your courses? (Pre-test)

2- During this semester you had different assessment activities including your midterm exams and classroom assessments. What kinds of assessment types were useful for you? (Post-test)

The ten students for the interview were chosen from the experimental group randomly to see what kind of assessment preferences and test anxieties they had at the beginning and what has changed at the end of this formative assessment implementation. At the beginning of the first interview, the students were first asked to introduce themselves shortly. They were ensured that their names would not be revealed at all. Then, they were asked how they were assessed in their high schools by their teachers. Besides, their favorite assessment types were asked to get an idea about their preferences. Additionally, the students were asked if they had any fears about the assessment of the present course.

The second interview began with similar questions such as introducing themselves. However, this time they were asked what different assessment types they had during this course period, and their ideas about what and why they liked most about those assessment activities. Lastly, they were asked to indicate if they still had any fears about the assessment of this course. The excerpts taken from students' interviews are not corrected.

3.3.4. Teacher Observations and Field Notes

Teacher observation is a method of monitoring student behaviors from direct observations. During formative assessment implementation, the teacher usually observed the students while they were busy with self and/or peer assessment activities, and it yielded a description and analysis of teacher's informal assessment practices in the class. It explored the extent to which teacher and pupils shared an understanding of the nature and purpose of assessment events. The data collected through teacher observation was recorded as the field notes of the researcher.

The field notes of the researcher provided valuable insights into conversations, emotions, and behaviors; interviews conducted at the beginning and end of the study may not have captured these specific occurrences. Taylor and Bogdan (1998) state that

"the field notes should include descriptions of people, events, and conversations as well as the observer's actions, feelings, and hunches or working hypothesis." (p. 67). Through observation, questioning, and reflections, field notes were a valid and important piece of this study. Observation and information about the formative assessment applications during the classroom sessions became the records within the researcher's field notes.

According to Fraenkel and Wallen (2006), field notes reflect the researcher's detailed notes that s/he takes as s/he observes what is going on in an educational setting (classroom or school). To ensure capturing various behaviors, and emotions of the students participated in this study, the researcher recorded his thoughts and subjective reflections after each classroom session. Besides, the key words, phrases, notes about events were recorded immediately after the classes.

3.4. Data Collection Procedure

The following parts will be explaining the steps of the present study, and it includes the procedures for data collection. The first section (pilot study, 3.4.1) will give the very first steps of data collection decisions and tool development stages. Also it will explain how and why the pilot study participants selected, and their study steps. Lastly, it will clarify what changes in participants and data collection tools have been made for the main study.

The second section (main study, 3.4.2) will begin with explaining the participants, and selected data collection tools after piloting. Next, the formative assessment applications will be illuminated through the field notes, which were developed while the chapters and topics of the course were being taught and covered during the classroom activities. And lastly, the post-test applications will be explained for data collection.

3.4.1. Pilot Study

A pilot study is a small scale examination intended to test data collection tools, and the design of the main study, and to improve the main study's quality and efficiency (Teijlingen & Hundley, 2001, Reiss, n.d). Also, a pilot study can reveal deficiencies in

the design of a proposed experiment or procedure and these can then be addressed before time and resources are expended on large scale studies (Cohen and et al, 2007).

The participants of the pilot study for the present research were 107 students of ELT department at Cukurova University. They were the first year students whose ages varied from 18 to 20, and they had the same university entrance exam. The study was conducted 2008-2009 educational year, previous year of the main study, in *Advanced Reading 1* course.

The focus of the pilot study was mainly on the formative assessment applications in the classroom. The aim for using checklist was that the use of these checklists keeps students on track and allows them to take responsibility for their own learning through peer- and self-evaluation. Also, the checklists clearly communicate performance expectations in terms of criteria and standards (Emery et al, 2006). The first step was to prepare and hand out checklist covering the topics of the week. In each week or for each chapter, students were given a checklist to let them see if they have any idea about that day's topic.

The second application was sharing learning goals of that week's topic or chapter. This helped them realize what they knew and what they need to know, which means the gap between pupils' current and desired level of learning or performance (Black and Wiliam, 1998a). This target sharing activity made the curriculum more transparent for students, and helped them to realize what was expected from them.

The third step in formative assessment application was the feedback, verbal and written. Black and Wiliam (1998a) consider assessment as formative when it uses feedback to adjust teaching and learning activities. Therefore, the instructor not only offered feedback to students for their learning, but he also got feedback from students to adjust his own teaching.

The next assessment application was to include students into their own assessment, and into peer assessment. Crooks (2001) states that that feedback on assessment cannot be effective if students do not accept that their work can be improved and recognize important features of their work that they wish to develop. In classroom

students were asked to evaluate their own work and correct their mistakes, mostly after a pop up quiz or an exercise.

The last application was the Test Anxiety Inventory and interview. This inventory application was useful both a couple of reasons; to see whether students had any difficulty in understanding the items, to analyze the reliability of the inventory, and to get an idea about students' test anxiety levels after formative assessment implementation. The interview was conducted with five students who were randomly selected, and the students were asked what they felt during the assessment of this course.

3.4.1.1. Shortcomings of the Pilot Study

As a result of the analysis of the pilot study data and formative assessment applications, some of the points were revised as:

a) The checklist usage for pilot study was very detailed and time consuming, therefore, the checklist prepared for the main study covered the chapter and its topic titles only.
b) During sharing the learning targets with students, it was realized that this activity was not enough for students to realize what they were really expected to know, so for the main study, this goal sharing activity was supported by both the given checklist to students and short discussions with students to internalize the target.
c) For a better formative assessment application, the verbal feedback to students was given more often via questioning techniques. Written feedbacks were also used more often; even the mid-term exam papers were included detailed feedback for students to what they knew and what else they need to learn. Also, the questioning helped the instructor to realize on what topics he had to work on more, and gave the chance to change some points in his teaching.
d) The pilot study included self and peer assessment activities, but it was found that they were not satisfactory for the students. For that reason, during the main study students were encouraged to assess their own and peers' work

more often, and the instructor was not any more doing the necessary evaluation on his own, but just helping the students assess their own and classmates' work. The students not only gave their peers verbal feedback but also wrote their comments on their peers' works.

e) Lastly, it was found that Test Anxiety Inventory was not enough to learn exactly what they feared, and their anxiety levels. Therefore, another tool, *Assessment Preference Scale,* was used to gather more data about not only the anxiety levels of the students but to see what they preferred before and after the formative assessment implementation. Moreover, the interview was conducted twice, before and after the study, to see the changes in students' anxiety levels and assessment preferences.

3.4.2. Main Study

The data from the experimental and control groups were collected through Test Anxiety Inventory (TAI) and Assessment Preference Scale (APS) at the beginning of the study and at the end of the semester. The experimental group students were first handed out TAI and then APS before the formative assessment implementation, and they were asked to fill in them at the end of the study again. The control group was applied the same procedure, but they did not get any kind of implementation during the semester. The data collected from the two groups were analyzed using SPSS program 13th edition. The data were analyzed by using "paired sample t-test" of this program.

The main purpose of the study was to see the effects of formative assessment on students' test anxiety and their assessment preferences. For this aim, the implementation of "Formative Assessment" in Contextual Grammar course for this study was conducted for one semester of an academic year. The program involved fourteen weeks, and the participants and the researcher met three hours each week. The main study will be explained through the topics covered during the fourteen week period.

Sentence Structure (Chapter 1):

At the beginning of the lesson a checklist (See appendix 4) that covers the topics of the first four weeks was handed out, and the students are asked to tick one of the

options such as "*Yes, Partly, or I have no idea*". Then, the target structures of the topic were discussed (goal setting) with the students, and they were informed about what they were expected to learn during the lesson. Additionally, the students were asked about the previous knowledge about that day's topic.

After goal setting, the students were handed out a multiple-choice test (See appendix 5) about the sentence structure and they were asked to self assess their answers at the end of the class. Next, the topic was presented with a power point presentation, and the students were asked several questions covering the topic (questioning). The questioning helped the instructor to learn what the students learned up to that time and what they needed to learn more. Lastly, the students were given feedback by the lecturer about the topic.

Cohesion (Chapter 2):

At the beginning of the class, the students were informed about general outlines of cohesion and what they were expected to learn in this lesson. Then, a hand-out including some explanations and questions about the topic was given to students, and the presentation began. The lecturer, from time to time, stopped the presentation and asked the students if they had any questions about the topic. Due to the fact that cohesion is a difficult topic for the freshman students, some sub-titles were (cohesion, parallelism) repeated with different examples. At the end of the lesson, the hand-out given at the beginning of the lesson was evaluated by the students. The students corrected their own mistakes, and they seemed happy during assessing their own answers…

The rest of the topics (/known-new contract) that were not presented in the first week were covered in the following week. Again, the students had some difficulties in understanding some of the topics, and they were given extra hand-outs and exercises to be completed in the class. Although this chapter was a really difficult one for the students, they were successful in completing the assessment tasks.

Sentence Rhythm (Chapter 3):

Firstly, the titles and the contents of the topic were explained at the beginning of the lesson, and it was observed that the students had no idea about the title *rhythm*. But it was understood via questioning that they partly knew about *what-cleft*. After presentation, the students were asked questions about the topic and gathered feedback about what they learned.

Writer's Voice (Chapter 4):

At the beginning of the class, the students were told that the titles of the day were *style, tone, diction*. When asked, the titles did not mean anything to the students. After presentation, they found out that they had partly knowledge about the topic. Through questioning and repeating some of the topics helped them learn more and feel more comfortable about the topic because at the beginning of the lesson they had fears about the mid-term exam if they would be asked about chapter 4. Lastly, they were given hand-outs and exercises covering the topics for self-assessment.

The Week after Mid-term Exam:

The students were asked if they wanted to see their papers or to solve the questions first. They wanted to see their papers and grades before solving the questions. After giving them some time to evaluate their exam papers, the lecturer began to solve the questions with students to help them understand their mistakes. Additionally, the answers were supported with a short video lesson about sentence structure taken from a web page. All the students stated that they would get a higher score because they knew all the answers. Solving all the questions was a way of repeating all the topics covered up to that time.

The students, moreover, were asked what could be done to understand the topics better, and they said that some more activities besides power-point presentation, hand-outs and the exercises of the course book would be helpful for them. The lecturer, after this feedback, offered them some videos about the topic for the following week. They seemed exited…

Coordination (Chapter 5):

Firstly, the lecturer gave some general information about coordination, and he shared the chapter goals with the students. Then, the students were asked several questions about the topic, and this questioning was very helpful for what else they needed to learn because it was better than asking them "Did you understand!"…

The presentation was supported with two short videos about the topic. They were about coordination, punctuation, run-on sentences. Meanwhile, the topic was not only presented with lecturer's explanations and power-point presentation, but it was also reinforced with questioning and the related videos. The students stated that the classroom sessions were more understandable and enjoyable in this way.

Subordination (Chapter 6):

At the beginning of the lesson, the students were asked about last week's topic (coordination). After getting satisfactory answers, the lecturer explained today's topic outlines and required targets with the students. Next, the power-point presentation and other materials about the topic were shared with the students. Then, the students watched the videos about subordination, and they were asked questions about the videos. Lastly, extra materials, such as exercises from other resources, were handed-out to the students.

Choosing Verbs (Chapter 7):

Before getting to this chapter, last week's topic (subordination) was overviewed with some questions. Then, the chapter, choosing verbs, goals were discussed and shared with the students. After that, the students were asked to write a long paragraph about the given topic (low-paid jobs), and they were told not to do anything on those paragraphs and focus on that week's subject. Next, the lecturer began to present the subject matters through power point presentation and videos from a web page.

After presentation, the students were asked to switch the paragraphs they had written at the beginning of the lesson with their peers, and the peers were required to

evaluate their friends' paragraphs considering this week's topic and find out the mistakes (peer assessment). It was observed that they seem really willing to do this task, and having fun while doing it… Finding out their peers' mistakes, without any anxiety of being evaluated for grades, both helped the ones who were evaluating and who were evaluated in creating awareness about their own paragraphs. Moreover, they seemed to understand the topics of the week better.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.0. Introduction

This chapter aims to present the analysis of the data gathered through the data collection tools: Test Anxiety Inventory, Assessment Preference Scale. The analysis will be done in two main sections: (1) the analysis and discussions of Test Anxiety Inventory, (2) and Assessment Preference Scale given both at the beginning and at the end of the implementation.

Relevant samples of interview will be given in each section of the analysis. The excerpts that are chosen as examples constitute representative samples for the rest of the group. As explained in chapter 3, the participants used English during the interviews, and their language mistakes were not corrected.

4.1. Test Anxiety Inventory

Test Anxiety Inventory will be analyzed in two sections. First, the overall results will be discussed between the experimental and control group. In other words, a comparison between the two groups as an overall result for TAI will be given in the following section, and statistically significant differences in the overall results will be analyzed accordingly. Besides, the total results within the experimental and control group will be analyzed; the change within the two groups at the end of the study will be examined. Second, seven categories of TAI (see section 3.3.1) will be analyzed separately, and analysis of each category will be done between and within the experimental and control group.

4.1.1. Overall Results of Test Anxiety Inventory Between and Within Experimental and Control Group

This section analyzes Test Anxiety Inventory (TAI) results first between the experimental and control group and them within the two groups. The mean scores of TAI between experimental and control group in Table 4.1. The pre-test results do not

show any statistically significant difference between the experimental and control group students in total (p= .674), which means the anxiety levels of the groups were similar at the beginning of the study (X_{exp}= 39.44, X_{cont}= 39.76). When we examine the post-test results, it can be seen that total means show statistically significant difference between the experimental and control group (p=.002). The experimental group's mean indicates a certain decrease (X_{pre}=39.44, X_{post}= 36.12) after formative assessment implementation whereas the control group's post-test mean remains almost the same as of the beginning of the study (X_{pre}= 39.76, X_{post}= 39.58).

Table 4.1. Overall Test Anxiety Levels Between The Groups

			Mean	**sd**	**p**
Pre-Test	Experimental	8	39.44	3.98	.674
	Control	8	39.76	3.02	
			X	**sd**	**p**
Post-Test	Experimental	1	36.12	4.03	.002*
	Control	1	39.58	4 .20	

When the results of TAI are analyzed within groups in Table 4.2, it is seen that the experimental group's mean (X_{pre}= 39.80, X_{post}= 36.12) decreased at a statistically significant level after formative assessment implementation (p= .003). On the other hand, the control group's result does not indicate a statistically significant difference (p= .858): the means of control group in pre and post-tests (X_{pre}= 39.97, X_{post}= 39.82) are almost the same at the end of the study.

Table 4.2. Overall Test Anxiety Levels Within the Groups

			Mean	**sd**	**p**
Experimental Group	Pre-test	1	39.80	3.96	.003*
	Post-test	1	36.12	4.03	
			X	**sd**	**p**
Control Group	Pre-test	5	39.97	3.31	.858
	Post-test	5	39.82	4.45	

From the results, it can be concluded that formative assessment application created a clear decrease in the experimental group's total anxiety level when compared with the control group's total anxiety level which remained same as of the beginning of the study. In other words, formative assessment implementation helped the experimental group students feel less anxious about the tests.

4.1.2. Results of Categories

Test Anxiety Inventory (TAI), as explained in section 3.3.1, has seven categories: *Others Views, Self-Image, Future Security, Bodily Reactions, Thought Disruptions, and General Test-Taking Anxiety*. In this section, each of the TAI categories is analyzed and discussed between and within the experimental and control group separately.

4.1.2.1. Category 1: Others' Views If They Do Poorly

In this category, the items are about the anxiety created by other people's views on students' test scores. The four items in this category aim to find out whether the behaviors and views of the people around the students create any anxiety on them. The items of this category are as follows:

- *People will question my ability if I do poorly.* (Item 9)
- *My friends will be disappointed in me if my score is low.* (item 21)

- *I am going to find out how others did before I announce my score.* (item 24)
- *Some people I know will be amused if I score low, and this bothers me.* (item 25)

Table 4.3. Others' Views About Them Between The Groups

			Mean	**sd**	**p**
Pre-Test	Experimental	8	5.02	1.07	.225
	Control	8	5.34	1.19	
			X	**sd**	**P**
Post-Test	Experimental	1	5.09	1.44	.807
	Control	1	5.16	1.18	

The means of the experimental and control group (X_{exp}= 5.02, X_{cont}= 5.34) in Table 4.3 do not show a statistically significant difference in the pre-test results (p= .225). When analyzed, the post-test results do not indicate a statistically significant difference (p= .654) between the means of the two groups (X_{exp}= 5.09, X_{cont}= 5.16).

Table 4.4 Others' Views about Them Within The Groups

		N	**X**	**sd**	**p**
Experimental Group	Pre-test	31	5.12	1.14	.929
	Post-test	31	5.09	1.44	
		N	**X**	**sd**	**p**
Control Group	Pre-test	45	5.26	1.17	.674
	Post-test	45	5.15	1.22	

When we analyze the results within the groups in Table 4.4, post-test results of the experimental group indicate a very slight decrease after formative assessment application (X_{pre}= 5.12; X_{post}= 5.09). The post-test results of the control group show a

similar situation (X_{pre}= 5.26, X_{post}= 15.15). However, this slight decrease in both groups do not show a statistically significant difference (p_{exp}= .929, p_{cont}= .674).

There may be some reasons why a significant difference in experimental group's anxiety level did not appear. Firstly, due to the past experiences, students probably still take other people's views serious. This anxiety was clear in the researcher's field notes. For example, even after the final exam, most students were still anxious about their families and friends' views about their test scores rather than considering what they need to learn to get a higher score next time.

The informal conversations with the students also show that other people's views are still important to them. People around them such as their friends and family members usually ask about their tests at school and comment on their test scores, which certainly create a tension and anxiety on students. These anxiety creator views were most effective during their preparation to university entrance exam the year they started to university. The students felt a lot of pressure to be successful and be a university student. This pressure is still effective on them because this is their first year at university and their friends and families wonder if they can be doing well as students at a university, which does not help them lessen that anxiety created by other people around them.

4.1.2.2. Category 2: Students' Self-Image

The second category is about the anxiety on students' self-image. These three items of *self-image* intend to discover if there is a change in their self-image anxieties or not. The items related to students' self-image are:

- *Getting a good score on one test does not seem to increase my confidence on other tests.* (Item 1)
- *If I do not do well on a test, I guess it will mean I am not as smart as I though I was.* (Item 15)
- *I do not feel confident and mentally relaxed before a test.* (Item 20)

Table 4.5. Students' Self Image Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	31	4.07	1.04	1.000
	Control	31	4.07	.88	
		N	**X**	**sd**	**P**
Post-Test	Experimental	38	4.06	.92	.716
	Control	38	4.16	.89	

The analysis between the means of the experimental and control groups (X_{exp}=4.07, X_{cont}= 4.07) in Table 4.5 does not indicate a statistically significant difference (p= 1.000) in their pre-test results. Similarly, the post-test results of the two groups (X_{exp}= 4.06, X_{cont}= 4.16) shows that there is not a statistically significant difference (p= .716) at the end of the study.

Table 4.6 Students' Self Image Within The Groups

		N	**Mean**	**sd**	**p**
Experimental Group	Pre-test	31	4.19	1.10	.630
	Post-test	31	4.06	.92	
		N	**X**	**sd**	**p**
Control Group	Pre-test	45	4.11	.88	.464
	Post-test	45	4.26	.86	

It can be seen in Table 4.6 that there is a decrease in experimental group means (X_{pre}= 4.19; X_{post}= 4.06) in their self image whereas the means of the control group (X_{pre}= 4.11; X_{post}= 4.26) increased at the end of the term: these changes in the groups do not show a significant difference for both groups (p_{exp}= .630; p_{cont}= .464).

There may be several reasons in experimental group's not having a statistically significant difference in their self-image at the end of the study. First of all, in order to have a positive self-image, students need to find things that they can feel positive about in their test scores. But instead, according to the field notes, the first thing that the students look for was to find out what their score was and what they did wrong.

Another reason may be that in dealing with test anxiety the students should really feel that they have studied all they can and that they are really prepared for the test. Initially this may mean that they have to study harder than they have ever studied before in order to feel positive about their preparation. But the experimental group students, in the informal conversations with the students and in the researcher's field notes, never found themselves fully prepared for the exams; similarly, even if they got a high score in an exam they did not feel secure for other exams.

4.1.2.3. Category 3: Students' Future Security

The third category focuses on the test anxiety related to students' future security. They generally link their future goals with the tests they have to take, which is partly true in that it is a common way to use tests not only for assessing the success of the students in the classroom but also for finding better jobs careers. The four items of category aim to learn the anxiety levels of the students related to their future security. The items are as follows:

- *People who do well on tests generally end up in better positions in life* (Item 5).
- *Knowing that my future depends in part on doing well on tests upsets me* (Item 8).
- *Tests make me wonder if I will ever reach my goals* (Item 13).
- *My test performance is directly connected to my future success and security* (Item 26).

Table 4.7. Students' Future Security Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	31	7.23	.81	.418
	Control	31	7.10	.76	
		N	**X**	**sd**	**P**
Post-Test	Experimental	38	5.83	.96	.455
	Control	38	6.00	.89	

The pre-test means of the experimental and control group (X_{exp}= 7.23, X_{cont}=7.10) shown in Table 4.7 do not show a statistically significant difference (p=.418). Although both groups show a decrease in their means in the post-test means (X_{exp}= 5.83, X_{cont}= 6.00), there is still not a statistically significant difference (p=.455) between the groups.

Table 4.8 Students' Future Security Within The Groups

		N	**Mean**	**sd**	**p**
Experimental Group	Pre-test	31	7.29	.969	.000
	Post-test	31	5.83	.739	
		N	**X**	**sd**	**p**
Control Group	Pre-test	45	7.13	.814	.000
	Post-test	45	6.00	1.04	

The analysis of the items related to this category shows that there is a statistically significant difference in the results of the experimental group's future security level (p= .000) (Table 4.8): the mean of the experimental group in the post test shows a noticeable decrease (X_{pre}= 7.29; X_{post}= 5.83) at the end of the study. Similarly, there is a decrease in control group's pre- and post-test results at the end of the term (X_{pre} = 7.13, X_{post} = 6.00), which also shows a statistically significant difference (p= .000).

Although the experimental group's pre- and post-test results indicate a higher decrease than the control group' results after formative assessment application, it is clear from Table 4.5 that both groups' anxiety level related to future security seems to decrease at a statistically significant level. One of the reasons for this decrease in both groups might be that these students most probably had a very high anxiety about their future before taking university entrance exam. After being successful in that exam and being a university student in the Faculty of Education, they probably felt safer about their future because they knew that they would be an English teacher after graduating from this faculty.

Additionally, these students most likely do not consider the tests in their courses at university as threatening as university entrance exam because they are not life changing exams. In other words, they have more chances to get higher scores in case of a failure in a course since they are not one-time shot exams. Moreover, the assessment systems of their courses give them more opportunities to express their ideas and knowledge in different ways whereas university entrance exam is a multiple-choice type exam. This situation is clear in one of the students' post-test interview when he was asked what different assessment tasks they had during the semester and how those assessments made them feel:

ST3: We used…like weekly quiz; we matched the words. It was very good because we learn how to match the words, close sentences each other. And I learned a lot of things from it because in multiple choices we only choose the answer, and have to find the correct answer only. We do not give anything; we do not focus on. But in these, we learn more than multiple choice.

In conclusion, the students in both groups felt less anxious about their future because they did not see the exams as life changing, one-time high stake tests anymore. Besides, they had different types of assessments to show their knowledge and ideas, which helped them feel safer for their future success.

4.1.2.4. Category 4: Students' Being Prepared for a Test

This category is about whether the students feel prepared for a test or not. Usually, almost all the test-takers have the feeling of having a lot more things to study before an exam, and this feeling increases the anxiety levels of the students. The three items of the fourth are as follows:

- *I never seem to be fully prepared to take tests* (Item 10).
- *My anxiety about tests makes me want to avoid preparing fully, and this just makes me more anxious* (Item 16).
- *One of my problems is 'not knowing exactly when I am prepared for a test* (Item 22)

Table 4.9. Being Prepared For A Test Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	31	5.42	.79	.422
	Control	31	5.52	.55	
		N	**X**	**sd**	**p**
Post-Test	Experimental	38	3.58	.71	.000
	Control	38	5.58	.62	

The analysis of the experimental and control groups' results (X_{exp}= 5.42, X_{cont}=5.52) shows that there is not a statistically significant difference between their pre-test means (p=. 422) (Table 4.9). However, the post-test results reveal that the means of the two groups (X_{exp}= 3.58, X_{cont}= 5.58) show a statistically significant difference (p= .000) at the end of the study. This significant difference is due to the decrease in the mean experimental group at the end of the study, which is discussed in Table 4.10 below.

Table 4.10. Being Prepared For A Test Within The Groups

		N	**Mean**	**sd**	**p**
Experimental Group	Pre-test	31	5.48	.769	.000
	Post-test	31	3.58	.719	
		N	**X**	**sd**	**p**
Control Group	Pre-test	45	5.60	.539	.838
	Post-test	45	5.57	.656	

The analysis of the experimental group students anxiety levels given in Table 4.10 indicates a certain decrease in their post-test results after formative assessment application (X_{pre}= 5.48; X_{post}= 3.58), and this difference shows a statistically significant difference (p= .000). On the other hand, control group's pre- and post-test results remain almost the same at the end of the term (X_{pre}=5.60, X_{post}= 5.57), which does not show a statistically significant difference (p= .838).

This significant decrease in the mean of experimental group about *students' being prepared for a test* may be based on some factors. Firstly, assessment tasks during

the lectures did not generally intend to grade students; instead, they aimed to help students' learning process. For this purpose, the experimental group students evaluated their own work (self assessment); and in addition, they had the chance to do peer assessment according to the target discussed and determined at the beginning of the lesson with the students. This target-shared self/peer assessment gave the students the chance to know about the assessment and it lessened the anxiety level of the experimental group students.

Another factor might be that the students got both written and verbal feedback from their friends and the teacher in their assessment process. Such feedback showed the students what they were expected in their assessment and what else they needed to learn. This can be seen in one of the participants post-test interview when he was asked whether the feedback offered to him by his friends and teacher were useful for him or not:

ST6: Yes actually, I get many feedbacks from my teacher, and my friends even. These feedbacks were useful for me because I knew that I won't have any problems about lesson, about the assessments you have given us. And so they were all useful for me to know what to study next time.

Meanwhile, students' anxiety about not being fully prepared for a test was almost removed by involving them in their own assessment procedure, and creating an awareness to make students understand that tests and/or assessment tasks do not necessarily aim to grade their studies, but instead to help them learn more and learn what they need to study to be more successful.

4.1.2.5. Category 5: Students' Bodily Reactions

The fifth category is about the bodily reactions of the students related to their test-taking anxiety. The test-takers sometimes have different bodily problems such as not eating or sleeping before an exam. The items related to this category are as follows:

- *I do not enjoy eating before or after an important test* (Item 3).
- *Having to face an important test disturbs my sleep* (Item 6).

- *I often find my fingers tapping or my legs jiggling while taking a test* (Item 17).

- *I often fell physically panicky when I have to take a really important test* (Item 23).

Table 4.11. Bodily Reactions Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	31	5.42	1.10	.547
	Control	31	5.57	1.19	
		N	**X**	**sd**	**P**
Post-Test	Experimental	38	5.48	1.15	.910
	Control	38	5.51	1.15	

It can be seen in Table 4.11 that the means of the experimental and control group (X_{exp}= 5.42, X_{cont}= 5.57) do not indicate a statistically significant difference in the pre-test results (p= .547). Likewise, the analysis of the post-test results shows that there is not a statistically significant difference (p= .910) between the groups (X_{exp}=5.48, X_{cont}= 5.51).

Table 4.12. Bodily Reactions Within The Groups

		N	**Mean**	**Sd**	**p**
Experimental Group	Pre-test	31	5.45	1.20	.906
	Post-test	31	5.48	1.15	
		N	**X**	**Sd**	**p**
Control Group	Pre-test	45	5.60	1.17	.867
	Post-test	45	5.55	1.09	

The experimental group pre- and post-test results related to their bodily reactions are almost the same (X_{pre}= 5.45; X_{post}= 5.48), which does not show a statistically significant difference (p= .906) (Table 4.12). Similarly, the control group's post-test

result is close to their pre-test means at the end of the term (X_{pre}= .5.60, X_{post}= 5.55). This decrease does not show a significant difference (p= .867).

4.1.2.6. Category 6: Students' Thought Disruptions

This category focuses on the test anxiety based on the thought disruptions of the students. The test-takers usually have negative thoughts before or during the exam. The items in this category aim to find out the anxiety levels of the students related to thought disruptions and they are as follows:

- *During a test, I sometimes find myself having trains of thought that have nothing to do with the test* (Item 2).
- *I mentally freeze up on important tests* (Item 11).
- *I sometimes seem to defeat myself (think negative thoughts) whole working on an important test* (Item 14).
- *When taking a test, my emotional feelings interfere with my concentration* (Item 18).
- *The harder I work on some test items, the more confused I get* (Item 19).
- *During tests, I sometimes get so nervous that I forget facts I really know* (Item 27).

Table 4.13. Thought Disruptions Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	31	8.94	1.01	.824
	Control	31	8.89	.95	
		N	**X**	**sd**	**P**
Post-Test	Experimental	38	8.58	1.62	.196
	Control	38	9.16	1.63	

The experimental and control groups' thought disruption means in the pre-test (X_{exp}= 8.94, X_{cont}= 8.89) does not indicate a statistically significant difference (p=.824) in Table 4.13. When analyzed, there is not a statistically significant difference (p=. 196) in the post-test means of the groups (X_{exp}=8.58, X_{cont}= 9.16).

Table 4.14. Thought Disruptions Within The Groups

		N	**Mean**	**Sd**	**p**
Experimental Group	Pre-test	31	9.00	.856	.219
	Post-test	31	8.58	1.62	
		N	**X**	**Sd**	**p**
Control Group	Pre-test	45	8.97	.96	.406
	Post-test	45	9.22	1.69	

When analyzed, there is a decrease in the means of experimental group students' thought disruptions (X_{pre}= 9.00, X_{post}= 8.58) after formative assessment implementation (Table 4.14). However, this change in experimental group mean does not show a statistically significant difference (p= .219). On the other hand, the control group's mean (X_{pre}= 8.87, X_{post}= 9.22) increased at the end of the term. Similarly, this change does not show a significant difference (p= .406).

Although the difference in the means of the experimental group students is not statistically significant, the decrease in their means might be considered positive compared to the increase in the means of the control group. The thought disruptions of the students probably lessened because of the fact that the students marked their own work (self assessment) and were encouraged to raise questions about the content and procedure of the assessment. In other words, the students were involved in the assessment process. Most probably, this helped them create a learning culture in the classroom as they started to help each other rather than compete with each other. Moreover, the students had the chance for peer assessment. Therefore, all these assessment activities inevitably created stress-free environment before and during the assessment period, and most of students' thought disruptions and anxiety level of them lessened. This was also clear in the interviews when they were asked how they were felt about the assessment tasks at the end of the semester. Following is an example about this issue:

ST7: … Peer assessment is good because when I show my writing to my friends, when they show my mistakes I will feel more comfortable because I can improve in a good way…

It is known that encouraging the students' superficial and rote learning and recalling of isolated items result in forgetting what has been learned soon. In this study, the students during formative assessment implementation were supported for deep learning by both classroom and group discussions, and their understanding of the topics were instantly assessed by questioning techniques by the teacher, and the students themselves (self assessment) and their peers (peer assessment). Additionally, during this period a more criterion-based approach was used than a normative one which emphasizes competition between students rather than personal improvement. In this learning and assessment friendly classroom environment, the students did not feel any competition with their classmates, and they mostly focused on their own learning, which helped lessening their test anxiety. They felt the ownership of their assessment because formative assessment was consistent with student goals, which was shared with students, and because they were involved in the assessment process, not just subjected to it.

Finally, during this assessment procedure, students were offered descriptive feedback both from the teacher and their classmates, which ended up in a more relaxed learning environment. The students began to feel free about learning from others' views about their work. Also, the necessary adjustment in teaching was done by the teacher after collecting feedback from the students. This assessment and adjustment procedure led students to understand the topics better and to lessen the test anxiety because they were aware of what was expected from them in the test.

4.1.2.7. Category 7: General Test-Taking Anxiety

The last category is about the general test-taking anxiety of the students. The items of this category are as follows:

- *It seems to me that test sessions should not be made the formal, tense situations they are* (Item 4).
- *If exams could be done away with, I think I would actually learn more from my courses* (Item 7).
- *I have a hollow, uneasy feeling before taking a test* (Item 12).

Table 4.15. General Test-Taking Anxiety Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	31	3.31	.57	.446
	Control	31	3.23	.43	
		N	**X**	**sd**	**p**
Post-Test	Experimental	38	4.47	.72	.075
	Control	38	4.10	.68	

The pre-test means of the experimental and control group (Xexp= 3.31, Xcont=3.23) does not show a statistically significant difference (p= .446) (Table 4.15). According to the analysis of the groups' results (Xexp= 4.47, Xcont=4.10), there is not a statistically significant difference in their post-test means (p=. 075). The changes in the means of the experimental and control group are discussed in Table 4.16 below.

Table 4.16. General Test-Taking Anxiety Within The Groups

		N	**Mean**	**Sd**	**p**
Experimental Group	Pre-test	31	3.25	.514	.000
	Post-test	31	4.74	.728	
		N	**X**	**Sd**	**p**
Control Group	Pre-test	45	3.28	.458	.000
	Post-test	45	4.04	.705	

The analysis of the experimental group pre- and post-test results in Table 4.16 indicate an increase at the end of the term (X_{pre}= 3.25; X_{post}= 4.74), which shows a statistically significant difference (p= .000). Similarly, there is an increase in control group's pre- and post-test results at the end of the term (X_{pre}= 3.28, X_{post}=4.04), and this increase shows a statistically significant difference (p= .000) as well.

The means of experimental and control group students show that their general test-taking anxiety levels increased. This increase in both groups' means may be explained with uncontrolled situations of the groups. Firstly, the students are in their first year at university, and they still need more time to get used to it. This general test-

taking anxiety can be seen in one of the students' post-test interview when he was asked if he had any fears about the assessment of the course:

ST1: …I do not have much fear. But sometimes I fear some of tests generally. Every student has general fear.

The following excerpt is from a student who claimed that she enjoyed the lesson and she would do personal development, but she cannot because of the exams:

ST7: To be honest our lesson was good. I usually take pleasure in your lesson. I usually love. I take enjoyment. I want to make some personal development but I cannot because of exams, homework. That's why; they are making me feel irritated. I cannot show my ability in the exams, it makes me feel depressed.

As it is clear from the excerpts, the students still have fears about the exams and do not feel comfortable about the exams in general. It will probably take more time for the students to feel less anxious about the tests.

4.2. Assessment Preference Scale

This section explains the analysis of the Assessment Preference Scale (see section 3.3.2). It consists of two main categories a) *traditional assessment preferences* and b) *formative assessment preferences*. Each category is analyzed between and within the experimental and control groups. Firstly, a comparison between and within the experimental and control groups' overall results for each main category are given in the next section. Then the sub-categories of the two main categories are analyzed separately within the two groups. The assessment task of each category was scored from one to five points. This means that one student could give 5 points to each task. For example, if a category includes three tasks, the highest score might be 15 at most. The analysis of the main and sub-categories are supported with the relevant excerpts taken from the pre- and post-test interviews.

4.2.1 Overall Results Between And Within The Experimental and Control Groups Related To Traditional Assessment Preferences

This part gives an overall result between the experimental and control group students' mean scores of their traditional assessment preferences. The maximum total score of the two groups related to traditional assessment preferences may be maximum 35 because this part includes seven items (matching, multiple-choice, open-ended and explanatory questions, rewriting, clause test, combining sentences), each of which may be 5 points maximum.

According to the pre-test means of the experimental and control groups (X_{exp}= 22.70, X_{cont}= 22.66), there was not a statistically significant difference between the two groups in total (p= .971) at the beginning of the study (Table 4.10). The pre-test results of the two groups show that the groups' traditional assessment preferences can be considered high. In other words, both groups highly prefer this type of assessment. This might be due to the fact that these students were almost always assessed via multiple-choice type exam or other traditional assessment types in their high schools.

Table 4.17. Traditional Assessment Preferences Between The Groups

PRE-TEST				**POST-TEST**		
Traditional Assessment	**Mean**		**P**	**Mean**		**P**
	Experimental	**Control**		**Experimental**	**Control**	
TOTAL	22.70	22.66	.971	21.68	22.68	326

Similar to pre-test results, the means of the two groups' post-test (X_{exp}= 21.68, X_{cont}= 22.68) did not show a statistically significant difference in their post-test results (p= .326) (Table 4.17). When examined, the post-test means are still high. It can be concluded that both group students still prefer traditional type of assessment at a high rate. Assessment types used during their high school years have an undeniable effect on these students' traditional assessment preferences, and as they are freshman students they need more time for renovation in their assessment preferences.

Table 4.18. Traditional Assessment Preferences Within The Groups

		N	X	sd	p
Experimental Group	Pre-test	24	22.25	4.32	.486
	Post-test	24	22.87		
		N	**X**	**sd**	**p**
Control Group	Pre-test	30	22.16	4.38	.537
	Post-test	30	21.16		

When compared within the groups in Table 4.18, the results revealed that both the experimental group (X_{pre}= 22.25, X_{post}= 22.87) and the control group mean (X_{pre}= 22.16, X_{post}= 21.16) did not show a statistically significant difference (p_{exp}= .486, p_{cont}= .537). To further analyze traditional assessment preferences of the students, the components of traditional assessment were analyzed separately: *selected-response preferences, limited-production preferences, and production task preferences*.

4.2.1.1. Selected-Response Assessment Task Preferences

Selected-response items generally assess receptive skills. In response to a prompt, students select an answer from given options. Typically, no new knowledge is constructed; students simply recall or recognize information required to select the appropriate response. The selected response assessment tasks used in this study include: *multiple choice tests and* matching. The maximum total score of this part may be maximum 10 because this part includes two tasks, each of which may be 5 at most.

Table 4.19 Selected-Response Assessment Task Preferences Between The Groups

		N	Mean	sd	p
Pre-Test	Experimental	36	7.05	1.11	.739
	Control	36	7.16	1.18	
		N	**X**	**sd**	**p**
Post-Test	Experimental	31	7.22	1.47	.455
	Control	31	7.48	1.02	

The analysis of the experimental and control groups' pre-test means (X_{exp}= 7.05, X_{cont}= 7.16) does not show a statistically significant difference (p= .739) as can be seen in Table 4.19. Likewise, according to the means of the groups in the post-test (X_{exp}= 7.22, X_{cont}= 7.48), there is not a statistically significant difference between the two groups.

Table 4.20. Selected-Response Assessment Task Preferences Within The Groups

		N	**X**	**sd**	**p**
Experimental Group	Pre-test	29	7.20	1,60	.807
	Post-test	29	7.27		
		N	**X**	**sd**	**p**
Control Group	Pre-test	45	7.13	2.08	.722
	Post-test	45	7.24		

When analyzed, there is not a statistically significant difference between the means of experimental group's (X_{pre}= 7.20, X_{post}= 7.27) pre- and post-test in their selected-response preferences (p= .807) in Table 4.20. Similarly, the means of control group (X_{pre}= 7.13, X_{post}= 7.24) in the pre- and post-test are not statistically significant (p= .722). From the means of both groups, it might be said that the students highly prefer multiple-choice and matching tasks even at the end of the semester.

What might cause this stability is that these freshmen students were selected for university by a university entrance exam that is totally based on selected response items, mostly multiple choice questions. Moreover, they probably studied for this exam more than 3-4 years time. Therefore, almost all students still find it easier to find or recall the correct answer from the given choices. The students' selected response preferences are also clear in their interview conducted both before and after the study. Followings are two excerpts from two of the students' pre-test interview:

ST4: In the high school, we have been tested by YDS exam. Our teachers gave us 100 questions. We solved them. They assessed us according to this exam. This exam was multiple choice. I want to be assessed by multiple choice.

ST6: I think it can be multiple choices. For example, vocabulary tests, structures, etc. having a, b, c, d, e…

As clear from this interview transcript, the students were assessed only by multiple choice exams, and this assessment was used for preparing the student for their university entrance exam. Another student stated similar ideas when he was asked about the changes in his ideas about type of the assessment tasks at the end of the formative assessment application:

ST2: I do not have much fears but multiple choice was easier than others.

Both the quantitative and qualitative data revealed that students, when in high school, were assessed mostly by multiple-choice test format for getting prepared for university entrance exam. Certainly, the reason for using such tasks was that all the students had to prepare for this high stake test. The gate-keeping role of this test gives no other choice to teachers but assessing their students in this way. This yields students' getting used to these selected response tasks; other types of assessments, which may result in a better understanding and learning, seem hard, and/or unnecessary for the students.

4.2.1.2. Limited-Production Preferences

Characteristically, limited-production tasks are narrowly conceived, limited in length. In response to a question, students produce an answer out of existing and new knowledge. The fact that students must construct new knowledge means that at least some of their thinking must be displayed. As opposed to selected response items, the teachers can assess the students' knowledge more deeply with limited-production tasks. The limited-production tasks of the study include: *Rewriting*, *Clause test*, and *Combining sentences*. The maximum total score of the two groups related to traditional assessment preferences may be maximum 15 because this part includes three tasks, each of which may be 5 at most.

Table 4.21. Limited-Production Task Preferences Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	33	9.39	1.95	.625
	Control	33	9.63	1.72	
		N	**X**	**sd**	**p**
Post-Test	Experimental	30	9.06	1.98	.631
	Control	30	8.80	2.24	

When analyzed, there is not a statistically significant difference (p= .625) in the means of the experimental and control group in the pre-test results (X_{exp}= 9.39, X_{cont}= 9.63) (Table 4.21). Similarly, the post-test results of the groups shows that there is not a significant difference (p= .631) in their post-test means (X_{exp}= 9.06, X_{cont}= 8.80). It can be concluded that the limited-production task preferences of the experimental and control groups did not show a statistically significant change at the end of the study.

Table 4.22. Limited-Production Task Preferences Within The Groups

		N	**X**	**sd**	**p**
Experimental Group	Pre-test	27	9.48	2.50	.595
	Post-test	27	9.22		
		N	**X**	**sd**	**p**
Control Group	Pre-test	43	9.02	3.33	.716
	Post-test	43	8.83		

The analysis of the experimental group' results as presented in Table 4.22 reveals that although there seems to be a slight decrease in the means of pre and post-test (X_{pre}= 9.48, X_{post}= 9.22), it does not show a statistically significant difference (p= .595). Similarly, there is not a statistically significant difference (p= .716) in the means of the control group (X_{pre}= 9.02, X_{post}= 8.83).

Like selected-response assessment tasks, students' preference of limited-production tasks can be considered high both at the beginning and at the end of the study. This is possibly because of the fact that some parts of the university entrance exam they took the previous year included clause tests and rewriting questions, but in

multiple choice format. Still, this shows that they are also used to these types of tasks, and they have the tendency to choose such tasks even at the end of the study.

4.2.1.3. Production Task Preferences

Production tasks assess students' productive skills in which they produce a considerable, concrete product that reveals their understanding or evaluating of certain concepts and skills. It is similar to limited-production tasks in that students are required to construct new knowledge and not just select a response. However, product assessments tasks typically are more substantial in depth and length, more broadly conceived, and allow more time between the presentation of the prompt and the student response than selected-response and limited-production items. The production tasks used in the study are as follows: *Open-ended and Explanatory questions*. The maximum total score of the two groups related to traditional assessment preferences may be maximum 10 because this part includes two items, each of which may be 5 at most.

Table 4.23. Production Assessment Task Preferences Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	24	6.08	1.74	.925
	Control	24	6.16	1.84	
		N	**X**	**sd**	**P**
Post-Test	Experimental	25	5.88	1.81	.274
	Control	25	6.44	1.44	

According to the pre-test results given in Table 4.23, there is not a statistically significant difference (p= .925) in the means of the experimental and control group (X_{exp}= 6.08, X_{cont}= 6.16). The post-test results reveal that the means of the two groups (X_{exp}= 5.88, X_{cont}= 6.44) are still close to each other and does not show a statistically significant difference (p= .274).

Table 4.24. Production Assessment Task Preferences Within The Groups

		N	**X**	**sd**	
Experimental Group	Pre-test	26	6.00	2.52	
	Post-test	26	6.30		939
		N	**X**	**sd**	
Control Group	Pre-test	30	6.13	2.12	
	Post-test	30	6.10		932

The analysis results related to production task preferences are presented in Table 4.24. Accordingly, the means of the experimental and control groups are close to each other at the beginning (X_{exp}= 6.00, X_{cont}= 6.13). Similarly, the means of control group are almost the same with the pre-test result (X_{pre} = 6.13, X_{post}= 6.10) at the end of the study: there is not a statistically significant difference in both groups' results (p_{exp}= .939, p_{cont}= .932).

As it is clear from the analysis, both groups prefer production tasks less than the other two tasks, selected-response and limited-production tasks both at the beginning and at the end of the study. This is possibly because the open-ended and explanatory questions require constructing new knowledge, not just selecting from given answers. This might be a new type of task for some of the students, which the students met at the university. As the students are expected to be more productive in this type of task, it seems that the expected renovation in students' preferences for production tasks will take more time as these freshman students still show resistance for the change as a consequence of high usage rates of selected-response items in their life changing university entrance exam.

4.2.2. Overall Results Between and Within The Experimental and Control Groups Related To Formative Assessment Preferences

This part gives an overall result between the experimental and control group students' mean scores of their formative assessment preferences. It is important to determine students' formative assessment preference because it is the new type of assessment implemented to the students of experimental group. Since formative assessment is not a test but a planned process involving a number of different activities,

during the course “Contextual Grammar”, the experimental group students were given the chance to participate in a number of formative assessment types.

Although students’ preferences of the formative assessment tasks below are analyzed and discussed separately later in this chapter, overall preferences of the students about formative assessment are outlined here. The maximum total score of the two groups related to formative assessment preferences may be maximum 35 because this part includes seven items (self/peer assessment, verbal/written feedback, weekly quiz, classroom activities, take home tasks), each of which may be 5.

Table 4.25. Formative Assessment Preferences Between The Groups

PRE-TEST				**POST-TEST**		
Formative Assessment	**Mean**		**P**	**Mean**		**P**
	Experimental	Control		Experimental	Control	
TOTAL	26.87	27.12	.857	30.60	25.95	.002

According to the pre-test means of the two groups (X_{exp}= 26.87, X_{cont}= 27.12), at the beginning of the study, there was not a statistically significant difference between the two groups in total (p= .857) in their formative assessment preferences. Contrary to the pre-test results presented in Table 4.25, that there is a statistically significant difference in experimental and control group means in the post test results (p= .002). While the experimental group shows a certain increase in their post-test means (X_{pre}= 26.87, X_{post}= 30.60), the control group shows a decrease (X_{pre}= 27.12, X_{post}= 25.95).

The decrease in the means of control group in Table 4.25 may be explained with measurement error. When the analysis was done between the experimental and control groups, the SPSS program accepted the number of the both groups as 16 for pre-test and 20 for post-test. However, the numbers were accepted as 17 for experimental group and 21 for control group when the analysis was conducted within the groups as can be seen in Table 4.26.

Table 4.26. Formative Assessment Preferences Within The Groups

		N	**X**	**sd**	**p**
Experimental Group	Pre-test	17	26.52	1.07	.004
	Post-test	17	30.11		
		N	**X**	**sd**	**p**
Control Group	Pre-test	21	25.76	1.12	1.000
	Post-test	21	25.76		

When the formative assessment preferences are analyzed within the groups in Table 4.26, the mean experimental group (X_{pre}= 26.52, X_{post}= 30.11) shows a sharp increase whereas the control group (X_{pre}= 25.76, X_{post}= 25.76) remains the same at the end of the study, and this increase in the mean of the experimental group is statistically significant (p= .004).

Although both groups had similar ideas at the beginning of the study, the experimental group students' preferences about formative assessment tasks moved to a higher level. In other words, their attitudes to this category developed in a better way after the formative assessment application. This high increase in experimental group students' formative assessment preferences may be due to several reasons. Firstly, one of the main activities of formative assessment was in the perception by the learner of a gap between the desired goal and his/her present state. This gap filling activity was provided by the instructor through communicating the targets of the topics of the day.

Another reason was the action by the learner to close that gap to attain the goal that the teacher shared at the beginning of the lesson. For this purpose, the key action was carried out by the students, which were marking their own work and raising questions about the assessment and the material covered by the assessment. This self assessment process created awareness in students' learning.

The third reason for this increase in experimental group students' formative assessment preferences was formative feedback. By direct and immediate feedback, not only the teacher got a view of both individual and class performances but the also students learned how well they did. The teacher could judge the students' success and

plan adjustments based on the formative feedback obtained from the students. Moreover, summarized formative feedback provided a basis for the teacher to reexamine topics in the unit and adjust his teaching if necessary. In addition to benefits of formative feedback for the teacher, students could also see progress in their learning. Namely, both teachers and students learned from the feedback results.

Since students became active in their own assessment, they were motivated to learn, which, in turn, resulted in a less threatening learning and assessment environment. This anxiety-free environment is also clear in students' interviews. Following excerpt taken from a student's first interview is an example related to this issue:

ST3: The first time I entered your course, I was not sure about how can I do that. I felt the first time this one. But the other classes like speaking I can do, but this one requires more energy than other ones. That's why I have fears. I want to be assessed by multiple choice. You know, our behaving in the class, attendance in the class, could be very nice.

The same student expressed how he got rid of this fear at the end of the study:

ST3: We used…like weekly quiz, we matched the words. It was very good because we learned how to match the words, close sentences each other. And I learned a lot of things from it because in multiple choices we only choose the answer. We do not give anything, we do not focus on. But in these, we learn more than multiple choice. Actually, I do not have any fears anymore, because, at the beginning, I did not know anything about the lesson. That's why; if I do not know anything about the lesson, I was afraid. Now I learned how to match, how to make sentences. I learned all of them; that's why, I do not afraid now.

As can be seen from the example above, the student had fears about both the course itself and its assessment at the beginning of the study, and preferred to be assessed by only multiple choice tests. After the application, he began to understand that there were much better assessment activities than multiple choices tests for their own learning. Similarly, another student stated in his second interview that all those

formative assessment activities were more enjoyable and useful for them because they could understand the lesson better:

ST6: Actually, with peer assessment, I can understand lesson better. It will be better if I continue like that. I like evaluating their situation and my situation. It improves me to know structures in language. To write good sentences, to combine sentences with conjunctions or coordinating conjunctions… It helps me do less mistakes…I get many feedbacks from my teacher, from my friends even. These feedbacks were useful for me because I knew that I won't have any problems about lesson, about the assessments you have given us. And so they were all useful for me…Actually there were many changes in my education, especially in this lesson. At the beginning of this semester, I had difficulty in understanding the lesson, I could not concentrate on the structures… but now I have less difficulty and I can write sentences better.

As we can understand from the statistically significant data given in Table 4.26 and the interviews, the students' preferences about the formative assessment seem to be very useful for them. Their assessment preferences moved from a multiple choice test format to a more learning-based assessment method, formative assessment. They clearly preferred this type of assessment tasks because they felt that they learned more and felt less frightened about the assessments. More specific analysis and discussions about three categories (*self/peer assessment, feedback, and ongoing assessment preferences)* related to formative assessment task preferences are done in the following sections.

4.2.2.1. Students' Self/Peer Assessment Preferences

In this study, self or peer assessment was the process of students' or their peers' assessing assignments. For self assessment, checklists were often used by the students. At the end of the lesson or a topic, the students were asked to evaluate what they had learned up to that point through the checklists they were given at the beginning of the lesson. For peer assessment, on the other hand, the students were asked to exchange their writings with their friends and to assess their friends. This part includes two assessment tasks: *self/peer assessment*. Therefore, the maximum total score of the two groups in this section may be maximum 10 at most.

Table 4.27. Self/Peer Assessment Preferences Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	28	6.07	1.69	.318
	Control	28	6.46	1.42	
		N	**X**	**sd**	**P**
Post-Test	Experimental	29	8.31	.92	.000
	Control	29	5.75	1.32	

The pre-test results given in Table 4.27 show that the means of the experimental and control group (X_{exp}= 6.07, X_{cont}= 6.46) do not show a statistically significant difference (p= .318). On the other hand, there is a statistically significant difference (p= .000) between the means of the two groups (X_{exp}= 8.31, X_{cont}= 5.75) in their post-test results.

Table 4.28. Self/Peer Assessment Preferences Within The Groups

		N	**X**	**sd**	**p**
Experimental Group	Pre-test	24	6.41	2.10	.000
	Post-test	24	8.33		
		N	**X**	**sd**	**p**
Control Group	Pre-test	40	6.25	2.04	.645
	Post-test	40	6.10		

The analysis of self/peer assessment preferences of the experimental group given in Table 4.28 reveals that there is a significant difference (p= .000) between the pre and post-test means (X_{pre}= 6.41, X_{post}= 8.33). The increase surely means that the students of experimental group found self and peer assessment useful and began to prefer more in the end. The means of the control group, on the other hand, (X_{pre}= 6.25, X_{post}= 6.10) do not show a statistically significant difference (p= .645).

This high increase in experimental group's self/peer assessment preferences may be due to in-class self/peer assessment activities. For example, after each topic was taught in the classroom, students were required to assess their own learning through some ways like checklists given at the beginning of the lesson or by evaluating their

own learning through exercises of the topic. The checklists covered the topics and goals of that lesson were given to the students at the beginning of the lesson.

The students were expected to evaluate what they knew about that day's topic. Then, at the end of the lesson they were asked to look at the same checklist to control what they learned and what could not understand. This self assessment process also helped the teacher to observe students' learning process because while the students were assessing their own learning, he had the chance to review their understanding individually.

Also, the students tried to assess their peers or classmates. This gave them a lot of opportunities to correct their own mistakes while trying to correct their peers'. At the beginning, the students were uninterested and unwilling for this practice because they were not sure about getting the ideas of their classmates or peers about their own work or learning, but during this period the students felt less threatened, and they became more involved and began to enjoy assessing their peer's learning. The following interview excerpts taken from two of the students in the second interview reveal this involvement process in students' self and peer assessment:

ST3: I believe peer assessment is very useful for me because I get knowledge from it. I let my friends assess me because I can learn my wrong usage of verb from them and he gives me how to correct them. And it was very good for me, and self assessment we made in the classroom was very good. Because we let ourselves control us and what we know. We see our knowledge, and we can use them in the writing. If we learn, we should use them in the writing. It will stay life long in our memory.

ST4: I think self assessment is useful for me because I can look my error, or peer assessment is useful too because I can look my friend's error or my friend can look my error. He says me and I correct my error.

As it is clear from the interview data, students find self and peer assessment useful because they helped them to raise their awareness on their own strengths and weaknesses. More to the point, the process of self/peer assessment yielded an increased engagement with learning and assessment process. They helped the students especially set goals, clarify objects, take responsibility for their own learning, and thus increased

students' self confidence. Through self and peer assessment, students were able to see the mistakes in their thinking and could correct them for future assignments. By assessing their own papers, students were better in understanding the assessment process and could recognize their own strengths and weakness. Students learned how to think while completing assignments.

4.2.2.2. Students' Feedback Preferences

Feedback is one of the key elements of formative assessment. Feedback is the information which can be written or oral and should focus on helping pinpoint areas of strengths and weaknesses. The feedback preferences of the study are as follows: *feedback, verbal feedback, written* feedback. The maximum total score of the experimental and control groups related to feedback assessment preferences may be maximum 15 because this part includes two items, each of which may be 5 at most.

Table 4.29. Feedback Preferences Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	21	11.23	2.44	.864
	Control	21	11.33	2.28	
		N	**X**	**sd**	**p**
Post-Test	Experimental	23	12.17	2.03	.243
	Control	23	11.43	1.90	

The analysis of the experimental and control groups' feedback preferences revealed in Table 4.29 does not indicate a statistically significant difference (p= .864) in the means of the groups (X_{exp}= 11.23, X_{cont}= 11.33) in their pre-test results. Similarly, there is not a statistically significant difference (p= .243) in the post-test means (X_{exp}= 12.17, X_{cont}= 11.43).

Table 4.30. Feedback Preferences Within The Groups

		N	**X**	**sd**	**p**
Experimental Group	Pre-test	21	10.71	2.57	.033
	Post-test	21	12.00		
		N	**X**	**sd**	**p**
Control Group	Pre-test	25	10.96	2.48	.873
	Post-test	25	11.04		

Like self/peer assessment preferences, the experimental group's feedback preference means (X_{pre}=10.71, X_{post}= 12.00) indicate a significant difference (p= .033) at the end of the study (Table 4.30). The means of the control group pre and post-test are almost same (X_{pre} = 10.96, X_{post}= 11.04) in both tests, and do not show a significant difference (p= .873). The sharp increase in the means of experimental group shows that they began to prefer getting different types of feedback from their teachers and classmates.

One of the possible reasons for this increase of the experimental group might be that students regularly got written and mostly verbal feedback from the instructor during classroom activities. This feedback about students' learning was always constructive and encouraging. After the students' benefits of such feedback, they began to prefer getting both written and verbal feedback for their work.

The feedback offered during classroom sessions was in two ways: external and internal feedback. The external feedback is the information about students' work provided both by their peers and instructor. This feedback provided additional information which helped students to reexamine their knowledge about the topics. Teacher feedback served as a reliable external reference point which evaluated their progress and their own internal goals. The information provided by the teacher was in a dialogue form rather than information transmission. This dialogue with the teacher helped students to develop their understanding of expectations and standards, and to correct their mistakes.

Another source of external feedback was the classmates of the students. Peer dialogue was beneficial for students in some ways. First of all, students who just learned something were generally better than the teacher in explaining it to their classmates in a language which was accessible. In other words, the students were sometimes more capable of explaining what needed to be corrected than the teacher. Second, peer discussion showed the students alternative perspectives and tactics on problems. Those alternative views enabled students to revise or reject their initial premise and construct new knowledge and meaning through negotiation. Thirdly, by commenting on their peer's work, students developed objectivity of judgement (about their work in relation to goals or standards) that could be transferred to the assessment of their own work. Lastly, peer discussion was motivating and easier for students to accept critiques of their work from peers than the teacher. In the second interview with one of the participants, for example, it can be seen that the students felt more comfortable during peer assessment that provided them feedback:

ST7: …Peer assessment is good because when I show my writing to my friends, when they show my mistakes I feel more comfortable because I can improve in a good way. That's why, peer assessment is really important…

Another student stated that he liked peer assessment because it helped them increase the communication among their classmates:

ST5: I liked peer assessment most because it is some kind of classroom activity. At lesson students do not talk much each other, and I think; it was useful for our class. And peer assessment improved our communication with our friends, also, help our teacher improved our learning.

Internal feedback, on the other hand, is the information provided by the students themselves. During classroom sessions, the students did not only get feedback from their instructor and their classmates, but they also got feedback from self assessment activities. This self-generated feedback led to a reinterpretation of the task and to adjust their internal goals or strategies. Following excerpt taken from a student in the second interview supports this issue:

ST3: ... and self assessment we made in the classroom was very good. Because we let ourselves control us, what we know. We see our knowledge, and we can use them in the writing. If we learn we should use them in the writing. It will stay life long in our memory.

As it is clear, the feedback offered from the teacher and the peers were useful for the students, and helped them to internalize the objectives identified by the teacher at the beginning of the lesson. Besides, the formative feedback was also useful for the teacher to do necessary adjustments in his teaching.

4.2.2.3. Students' Ongoing Assessment Preferences

In this study, ongoing assessment was carried out on a repeated and periodic basis throughout the semester. Besides self and peer assessment tasks as the two of the main tools for ongoing assessment, different classroom procedures were applied to students: *weekly quiz, classroom activities, and take-home* tasks. The maximum total score of the two groups related to ongoing assessment preferences may be maximum 15 because this part includes three items, each of which may be 5 at most.

Table 4.31. Ongoing Assessment Task Preferences Between The Groups

		N	**Mean**	**sd**	**p**
Pre-Test	Experimental	34	9.26	2.04	.383
	Control	34	8.67	2.84	
		N	**X**	**sd**	**p**
Post-Test	Experimental	28	10.00	1.96	.025
	Control	28	8.64	2.24	

When analyzed, the pre-test results do not show a statistically significant difference (p= .383) between the experimental and control group means (X_{exp}= 12.17, X_{cont}= 11.43) (Table 4.31). The post-test results, on the other hand, reveals that there is a statistically significant difference (p= .025) in the means of the groups (X_{exp}= 10.00, X_{cont}= 8.64).

Table 4.32. Ongoing Assessment Task Preferences Within The Groups

		N	**X**	**sd**	**p**
Experimental Group	Pre-test	28	9.50	3.04	.393
	Post-test	28	10,00		
		N	**X**	**sd**	**p**
Control Group	Pre-test	43	8.75	3.19	.450
	Post-test	43	8.93		

As can be seen from Table 4.32, both the control group (X_{pre}= 8.7, X_{post}= 8.93) and the experimental groups' means (X_{pre}= 9.50, X_{post}= 10.00) about ongoing assessment preferences show an increase. However, this increase does not indicate a significant difference (p_{cont}= .450, p_{exp}= .393). It might be said that the experimental group students preferred ongoing assessment tasks such as weekly quiz and classroom activities more than the control group at the end of the study.

Although there is not a significant increase in the experimental group's result about ongoing assessment preferences, the increase can be considered positive. One of the possible reasons for this increase might be that in classroom sessions, students of the experimental group had quizzes regularly during the formative assessment application, which aimed to assess their learning processes. These quizzes were generally evaluated by students themselves and/or their peers. Those activities gave them the chance to see what they learned up to that point and what else they needed to learn.

Also, classroom activities in the classroom were undoubtedly taking place all the time. These activities were distinct from observed communicative activities in that they were always informal. And as the teacher emphasized the performance rather than simply knowledge, the students were more comfortable about these assessment activities. This situation can be seen in one of the students' post-test interview:

ST3: We used…like weekly quiz, we matched the words. It was very good because we learn how to match the words, close sentences each other. And I learned a lot of things from it because in multiple choices we only choose the answer. We do not give anything, we do not focus on. But in these, we learn more than multiple choice.

Although the students' mean results do not demonstrate a statistically significant difference, the experimental group students felt more willing about the ongoing assessment tasks such as weekly quiz, and classroom activities. They understood that the multiple choice tests are not the only way to see what they had learned. Moreover, the ongoing assessment tasks helped them internalize and realize what they really learned and what else they needed to acquire to be more sufficient about the topics.

CHAPTER 5

CONCLUSIONS

5.0. Introduction

The aim of this chapter is to present a summary of the findings about the present study and connect them to the research questions started up with. Moreover, this chapter compares and matches the findings with the related literature whenever possible. Implications of this study and further comments and recommendations as well as our reflections are touched upon in this chapter.

5.1. Review of Results - Research Question 1

One of the purposes of this study was to find out some possible effects of formative assessment on students' test anxiety. The research question for this purpose was:

1. Does formative assessment have any effects on students' test anxiety?

To carry out this aim, Test Anxiety Inventory (TAI) was applied to both experimental and control group students. As mentioned in chapter 3 (see 3.3.1), this inventory had seven categories about test anxiety. The results of TAI were given as the overall results and the categories separately.

The overall results revealed that formative assessment crated a decrease in experimental group students' total *test anxiety levels*. This result is parallel with the findings of Svihla (2006) who reported in her study that students felt less anxious after formative assessment implementation. When the categories of TAI were analyzed, the first category that indicated a decrease in students' test anxiety was *future security*. One of the reasons that caused this encouraging decrease was the self/peer assessment tasks in the classroom. To be precise, students evaluated their own work and were encouraged to raise questions about the assessment, which built non-threatening assessment environment for the students. Peer assessment had similar effects on students' fears and helped them in relieving test anxiety about their future security.

Similarly, Black and et al. (2003) point that peer-assessment has been found to improve student motivation to work carefully. It also helped identify the strengths of the students and might help boost confidence. Therefore, the self/peer assessment gave the students their own share in classroom assessment during this study, which provided them a more secure assessment environment because the teacher was been the only one responsible for student assessment anymore.

Lastly, teacher using formative assessment changed the culture of their classroom, putting the emphasis on students' feeling safe to take risks and make mistakes and to develop self-confidence in the classroom. After formative assessment implementation, students began to know what they would learn and how they would be assessed, which made them more comfortable about their future studies and tests.

Another significant difference in the TAI categories was found to be in *being prepared for a test*. The students were anxious about getting prepared for test at the beginning of the study. However, their anxiety related to not being prepared for a test lessened sharply at the end of the study. This anxiety decrease was due to some factors. To begin with, most of the formative assessment tasks such as self/peer assessment were not strictly time-limited and formal. Due to the fact that some students had some problems with time limitation, this student-friendly assessment tasks reduced students' test anxiety before and after the test.

Another reason for this decrease in students' feeling not prepared for a test was that assessment tasks did not generally aim to score students' exams, but instead, they intended to help students' learning process. Meanwhile, students' anxiety about not being fully prepared for a test was almost got rid of by involving them to their own assessment procedure. This involvement created awareness in students' understanding that tests and/or assessment tasks do not necessarily aim to grade their studies, but to help them learn what they need to add to themselves to be more successful.

The next change found in the analysis of assessment preference scale was in students' *thought disruptions*. Although this chance in the experimental group students did not show a statistically significant decrease in this category, the decrease can be considered as positive. One of the factors causing this positive change in students'

thought disruptions was that the students marked their own work and were encouraged to raise questions about the assessment and the material covered by the assessment. This involvement in their own assessment process of the students was an essential aspect of reflection on their own learning, and it lessened possible thought disruptions related to their tests and their results. Moreover, they did peer-assessment tasks during the study. In other words, the students felt comfortable about the results of the test because their scores became a tool for further learning during this study, not a way of grading.

Finally, during formative assessment period a more criterion-based approach based on collaboration of students was used than a normative one which emphasizes competition between students. Ross (2005) states that "assessment incidents are not considered punctual summations of learning success or failure as much as an on-going formation of cumulative confidence, awareness, and self-realization learners may gain in their collaborative engagements with tasks." (319). The students did not feel any competition with their classmates, and they mostly focused on their own learning, which helped the students lessen their test anxiety. They felt the ownership of their assessment because formative assessment was consistent with student goals shared with students, and because they were involved in the assessment process, not just subjected to it.

Last category to be mentioned here is *general test-taking anxiety*. Both the experimental and control group analysis of post-test showed an increase at the end of the study. This increase may be explained with different reasons. Firstly, this was students' first year at the university, and they still needed more time to get used to the new types of the courses and assessments. Another reason was that the previous test experiences of the students might cause this general fear about the exams and they did not feel usually comfortable about the exams in general.

Other categories- others' views and self-image- that did not show any significant change may be due to some factors. In the first category, *others' views,* due to the past experiences, other people's views are still clearly important for the students. The informal conversations with the students showed that their families, friends' asking about their exams and the scores they get from the exams certainly creates a tension and anxiety on students. The second category, *self-image,* also did not indicate a significant difference. There may be some reasons for this stability. Firstly according to the

researcher's field notes, the only thing that most of the students were looking for was their test scores and what they did wrong in the exam, which surely did not help them feel positive about themselves and their tests. Another reason creating anxiety might be that the students need to feel that they have studied all they can for the test. However, researcher's field notes revealed that they never seemed to be ready for a test even if they did their best before the tests and even if they got high scores from other tests.

5.2. Review of Results – Research Question 2

Another of aim of this study was to find out the changes in students' assessment preferences. The research question for this purpose was:

2. Do students' assessment preferences change when they experience a formative mode of assessment?

To carry out this aim, Assessment Preference Scale was applied to both experimental and control group students. As mentioned in chapter 3 (see 3.3.2), this scale had 16 items scored on a five point Likert scale. It consists of two main categories a) *traditional assessment preferences* and b) *formative assessment preferences*.

From the analysis of the traditional assessment preferences of the students, it can be concluded that both the experimental and control group students still prefer traditional type of assessment at a high rate. This is because of undeniable effect of the assessment tasks used during their high school years, and as they are freshman students they need more time for renovation in their assessment preferences. Another explanation for preferring this type of assessment was that the students had to take a university entrance exam, which is mostly based on traditional assessment and does not directly measure the ability to produce language.

Certainly, the reason for high school teaches' using multiple choice tasks in their classroom assessment was that all the students had to prepare for this university entrance exam. As Stoneman (2005) states that these kinds of test results are still the main criteria for students' promotion or being accepted to educational institutions. For this reason, the gate-keeping role of this test gives no other choice to teachers for

assessing their students, and this type of assessment yields students getting used to this selected response tasks. Then again, other types of assessments, which may result in a better understanding and learning, seem hard and unnecessary for the students.

Additionally, the students studied for this exam more than three or four years, which means that all they were directed to choose from the given options for 3-4 years. Although students find some other assessment types useful for their personal achievement, they still showed a tendency for traditional type of assessment. This situation is also stated by Black and Wiliam (1998a), "they spend time and energy looking for clues to the *right answer*" (p: 143). Namely, as a result of this powerful exam, instead of producing any kind of language, the students prefer to be assessed through multiple-choice tests highly even after this formative assessment practice period.

The analysis of the *formative assessment preferences* of the experimental group students indicated that their preferences had a lot of gains about the formative mode of assessment. In other words, they added new types of assessment to their preferences. This change had some causes: firstly, one of the main activities of formative assessment was sharing the goals with the learners, which aimed to closing the gap between the desired goal and the students' present state. This gap filling activity was provided by the instructor through communicating the targets of the topics of the day. Similar to the findings Lynch and Maclean's in their study (2003), the present study found that the goal setting and target sharing activity created awareness in students' learning and assessments.

Another reason for this increase in experimental group students' formative assessment preferences was the feedback. According to Ramaprasad (1983), feedback is 'information about the gap between the actual level and reference level of a system parameter which is used to alter the gap in some way' (p. 4.; in Sadler, 1989, p. 120). The way for the students to close the gap during this study was the feedback from their peers and instructor, which enabled them to revise or reject their first idea and construct new knowledge and meaning through negotiation.

The present study's findings echo in Narciss and Huth's (2004) study where they put forward that systematic feedback had positive effects on students' motivation. Moreover, summarized formative feedback provided a basis for the teacher to reexamine topics in the unit and adjust his teaching if necessary. To be precise, both teachers and students learned from the feedback results.

From the analysis of assessment preference scale, another significant change was found to be in students' *self/peer assessment* preferences. As Black and et al. (2003) pointed, self/peer assessment is exclusively important in students' learning; these types of assessment were found to improve student motivation to work carefully. In line with their study, the self/peer assessment activities during formative assessment implementation in this study had a positive effect on students' assessment preferences because they found this type of assessment useful:

ST1: I think self assessment is useful for me because I can look my error; or peer assessment is useful too because I can look my friend's error or my friend can look my error. He says me, and I correct my error.

To sum up, the process of self/peer assessment resulted in an increased engagement with learning, and it also increased students' confidence. During self and peer assessment, students were able to get a view of their thinking and mistakes; then, they could correct them for future assignments. By assessing their friends and own papers, students were better in understanding the assessment process and could recognize their own strengths and weakness.

5.3. Suggestions for Curriculum and Teaching

As formative assessment has been accepted and proved to be a very powerful way of enhancing student learning through various assessment tasks (Black & Wiliam, 1998a; Black & Wiliam 1998b; Brookhart, 2007; Butler & McNunn, 2006), in an effort to distill the quality of classroom assessment, students' learning, teachers' professional improvements, the following statements are my personal suggestions derived from the current research for a better student learning:

- There should be a link between curriculum, instruction and assessment, which is clearly understood by the students.
- Teachers should have a consistent and ongoing plan for their professional development.
- Teachers should share the learning goals with students, and discuss the assessment questions with them critically, so there will be more reflection on what is being assessed.
- Instead of over-stressing the grading function, teachers should focus more on the learning functions of assessment.
- For an assessment to be formative, the assessment should be used as feedback for closing the gap between what the students know and what else they need to learn.

As Black and William (1998a) characterize assessment broadly to include all activities that teachers and students carry out to get information used diagnostically to alter teaching and learning, teachers can create many opportunities to assess how students are learning, and then use this information to make constructive changes in instruction. Meanwhile, if teachers do no use tests and assessment tasks to give feedback about learning, they are just the indicators of a final summative test.

5.4. Suggestions for Subsequent Research

As I reflected before in chapters one, three and four, different ways for further research become obvious. These ways can be gathered under different parts. Firstly, this study sought the possible effects of formative assessment on test anxiety. In this study, formative assessment was found to have positive effects in decreasing test anxiety of the students. For further research, these positive effects of formative can be the main focus of the future studies related to test anxiety.

Secondly, the limitation of this study was that it was conducted in only one course while formative assessment could be more useful in reducing test anxiety and increasing learning if it was applied in coordination with other courses and lecturers. That is, further researches may be conducted with other courses and lecturers in

coordination at the same time, which may give a clearer picture of formative assessment.

Another suggestion might be that although this study included two groups consisting of more than eighty students, larger size of a research population possibly will be more effective in understanding what formative assessment may add to students' learning. Next suggestion is that the current study has been conducted with freshman students at university. This age group showed a resistance in changing their assessment preferences because of the fact that this was their first year at university and their acquired test habits, mostly selected-response assessment tasks, had a very slow changing process. And even at the end of the formative assessment implementation, the students still wanted multiple-choice tasks in their exams despite adding new ways such as self/peer assessment to their preferences. Consequently, a follow up study may possibly be useful for seeing the effects of formative assessment on students' test anxieties and assessment preferences.

5.5. Personal Reflections

The findings of this study coincide with other research dating back to 1970s. High-stake and locally-mandated test scores have been used to measure the success of the students and teachers. This study did not question the role of high-stake tests in student learning. However, it revealed the effects and pressure of those tests in students test anxiety and assessment preferences.

In chapter two, formative assessment was defined by Black and Wiliam (1998a) as a powerful way to improve student learning — but summative assessments such as standardized exams can have a harmful effect. According to the current study data analysis, the role of assessment in English Language Teaching (ELT) continues to be a scoring tool rather that a way used for enhancing student learning. The current evidence shows that teachers' using formative assessment in the evaluation of the students is essential in order to improve student learning. Furthermore, ELT department students are the future English teachers. In other words, if we expect our students to improve their own students' learning we, as lecturers, should use formative assessment more often.

REFERENCES

Allal, L. & Ducrey, G. P. (2000), "Assessment of-or in- the zone of proximal development", *Learning and Instruction,* 10, 137-152.

Alotaiby, F. T. & Chen, J. X. (2005), *Generic Summative Assessment Functional Model*, Proceedings of the Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Networks (SNPD/SAWN'05).

Angelo, T. A. (1991), "Ten easy pieces: Assessing higher leaning in four dimensions", T. A. Angelo (Ed.) *Classroom Research: Early lessons from success*, 17-31. San Francisco, CA: Jossey-Bass, Inc.

Assessment Reform Group (2002), *Assessment for Learning*: *10 research-based principles to guide classroom practice*, Retrieved November 7, 2009 from http://www.assessment-reform-group.org/CIE3.PDF

Bachman. L.F. & Palmer A.S. (1996), *Language Testing in Practice*, Oxford University Press.

Baeten, M., Dochy, F. Struyven, K. (2008), "Students' Approaches To Learning An Assessment Preferences In A Portfolio-Based Learning Environment", *Instr. Sci. 36*. 359-374.

Bartram, B., Bailey, C. (2010), "Assessment preferences: a comparison of UK/international students at an English university", *Research in Post-Compulsory Education, 15/2*. 177-187.

Biggs, J. (2003), *Teaching for Quality Learning at University* (pp. 165-203), Buckingham, UK: SRHE and Open University Press.

Birenbaum, M. (1997), "Assessment preferences and their relationship to learning strategies and orientations", *Higher Education, 33*. 71–84

Birenbaum, M. (2007), "Assessment and Instruction Preferences and Their relationship with Test Anxiety and Learning Strategies", *Higher Education,* 53. 749-768.

Black, P. & Wiliam, D. (1998a), "Inside the Black Box-Raising the Standards Through Classroom Assessment", *Phi Delta Kappan. 80/2*. 139-148.

Black, P. & Wiliam, D. (1998b), "Assessment and Classroom Learning", *Assessment in Education: Principles, Policy & Practice, 5/1*. 7-74.

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003), *Assessment for Learning: putting it into practice*, New York, NY: McGraw-Hill.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004), "Working Inside the Black Box: Assessment for Learning in the Classroom", *Phi Delta Kappan, 86/1*. 8-21.

Boud, D. (2000), "Sustainable Assessment: rethinking assessment for the learning society", *Studies in Continuing Education,* 22/2.

Bresciani, M. J. (2007), *What Exactly is Assessment? What is your Role in It?,* (Online Document) Retrieved April 13, 2010 from http://www.ncsu.edu/assessment/presentations/uapr/what_exactly_%20is_assessment.pdf

Brookhart, S. M. (2001), "Successful Students' Formative and Summative Uses of Assessment Information", *Assessment in Education*, *8/2,* 153-169.

Brookhart, S. M. (2007), "Expanding Views about Formative Classroom Assessment: A Review of the Literature", J. H. McMillan. (2007), *Formative Classroom Assessment: From Theory to Practice*, *116-136*. Teachers College Press. New York.

Brookhart, S. M. (2008), *How to Give Feedback to Your Students*, United States. Association for Supervision and Curriculum Development. Retrieved December, 04, 2009 from http://www.ascd.org/publications/books/108019.aspx

Brown, H. D. (2004), *Language Assessment-Principles and Classroom Practices*, New York. Pearson Education, Inc.

Bryman, A. (2008), *Social Research Methods*, Oxford University Press, New York.

Burns, M. (2005), "Looking at How Students Reason", *Educational Leadership*, *63/3,* 26–31.

Butler, M. S. & McMunn, N. D. (2006), *A Teacher's Guide to Classroom Assessment: Understanding and Using Assessment to Improve Student Learning,* Jossey-Bass, San Francisco.

Cassady, J. C. & Gridley, B. E. (2005), "The Effects of Online Formative and Summative Assessment on Test Anxiety and Performance", *The Journal*

of Technology, Learning, and Assessment, 4/1. October. Retrieved December 12, 2009 from http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1059&context=jtla

Cederman, L. E. & Daase, C. (2003), "Endogenizing Corporate Identities: The Next Step in Constructivist IR Theory", *European Journal of International Relations*. *9/1*, 5-35.

Clapham, C. (2000), "Assessment and Testing", *Annual Review of Applied Linguistics,* 20, 147-161.

Cole, M. & Cole, S. (2001), *The Development of Children*, 4th Ed. New York: Scientific American Books, Distributed by W.N. Freeman and Company.

Cohen, L., Manion, L. & Morrison, K. (2007), *Research Methods in Education* (6th edition), Routledge, Oxon.

Coombs, S. & Fletcher, S. (2005), "Design and Evaluation of critical thinking scaffolds to support action research for teacher professional development", *Paper presented at the EARLI 2005 11th Biennial Conference*, University of Cyprus 23-25 August.

Cowie, B., & Bell, B. (1999), "A model of formative assessment in science education", *Assessment in Education*, *6*. 101-116.

Cresswell, J. W. (2003), *Research Design: Qualitative, quantitative, and mixed method approaches*, 2nd edition. Sage Publications, USA.

Crooks, T. (2001), "The Validity of Formative Assessment", *Paper presented to the British Educational Research Association Annual Conference*, University of Leeds, 13-15 September. Retrieved November 19, 2009 from http://www.leeds.ac.uk/educol/documents/00001862.htm

Derrich, J. & Ecclestone, K. (2006), "Formative Assessment in Adult Literacy, Language and Numeracy Programmes", *A Literature Review for the OECD*, Draft.

Dewey, J. (1944), *Democracy and education*, New York: Macmillan, Retrieved August 12, 2008 from http://www.ilt.columbia.edu/publications/dewey.html. [Originally published in 1916].

Dewey, J. (1998), *How we think* (Rev. Ed.), Boston, MA: Houghton Mifflin Company, [Originally published in 1933].

Donaghue, H. (2003), An Instrument to Elicit Teachers' Beliefs and Assumptions, *ELT Journal*, 54/7, 344-351.

D'Souza, L. A. (2009), "Assessing What Counts: Learning to Teach for Pupil Learning", *Unpublished Doctoral Dissertation*, Boston College.

Earl, L. (2003), *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*, Thousand Oaks, CA, Corwin Press.

Emery, L. J., Harvey, H. & Andersen, M. C. (2006), "Formative Evaluation Using Checklists to Improve Research Papers", *Perspectives in Health Informative Management,* 3/2, Retrieved April 13, 2010 from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2047304/

Erwin, T.D., 1991, *Assessing Student Learning and Development*, Jossey-Bass, Retrieved October 17, 2008 from http://www.qualityresearchinternational.com/glossary/assessmentoflearning.htm

Fisher, D. & Frey, N. (2007), *Checking for Understanding*, Virginia. Association for Supervision and Curriculum Development.

Forbes, E. W. (2007), "Improving the Knowledge and Use of Formative Assessment: A Case Study of A Model of Formative Assessment in A K-3 Science Curriculum", *Unpublished Doctoral Dissertation*, UMI.

Fraenkel, J. R. & Wallen, N. E. (2006), *How To Design And Evaluate Research In Education*, McGraw-Hill. NewYork.

Fulcher, G. & Davidson, F. (2007), *Language Testing and Assessment,* Oxon; Taylor and Francis Group.

Giebelhaus, C. & Bowman, C. (2002).Teaching Mentors: Is It Worth the Effort? *Journal of Educational Research*, *95/4*, 246-254.

Gijbels, D., Segers, M., Struyf, E. (2008), Constructivist Learning Environments And The (Im)Possibility To Change Students' Perceptions of Assessment Demands And Approaches To Learning, *Instructional Science,* 36, 431-443.

Hancock, C. R. (1994), Alternative Assessment and Second Language Study: What and Why?, *Eric Digest,* Retrieved August 18, 2008 from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/13/6f/18.pdf

Hangstrom, F. (2005), Formative Learning, *Communication Disorders Quarterly,* 28/1, 24-36.

Helle, H. J. (1988), *Soziologie und Erkantnistheories bei George Simmel*, Darmstadt: Wissesnschaftlische Buchsgesellschaft.

Harlen, W. (1998), "*Classroom Assessment: A Dimension of Purposes and Procedures", Paper presented at the Annual Conference of the NZARE*, Dunedin, December.

Harlen, W. (2005), "Teachers' Summative Practices and Assessment for Learning-tensions and synergies", *The Curriculum Journal.* 16/2, 207-223.

Harlen, W. (2007). "Formative Classroom Assessment in Science and Mathematics", J. H. McMillan. (2007), *Formative Classroom Assessment: From Theory to Practice*, 116-136, Teachers College Press, New York.

Hopkins, W. G. (2000), "Quantitative Research Design", *Sportscience,* 4/1, Retrieved March 23, 2010 from http://www.sportsci.org/2000/1/index.html.

Hopper, T. F. (2000), "Personal Construct Psychology for Developing Reflective Teaching in Physical Education: A Story of Decentering 'Self' as Teacher", *Avante,* 6/3, 46-56.

Institute of HeartMath. (2007), *Reducing Test Anxiety and Improving Test Performance in America's Schools*, Retrieved March 31, 2010 from http://www.heartmath.org/templates/ihm/section_includes/education/pdf/Tends_Summary_Results.pdf

Joia, L. A. (2002), "Assessment of a socio-constructivist model for teacher training: A case study", *Education Policy Analysis Archives*, 10(44), Retrieved June 7, 2008 from http://epaa.asu.edu/epaa/v10n44/.

Kelly, G. (1955), *The Psychology of Personal Constructs*, New York, Norton.

Klecker, B. M. (2007), "The Impact of Formative Feedback on Student Learning in an Online Classroom", *Journal of Instructional Psychology,* September, Retrieved November 6, 2009 from http://findarticles.com/p/articles/mi_m0FCG/is_3_34/ai_n27416665/?tag=content;col1

Kreber, C., Castleden, H., Erfani, N., Lim, J. & Wright, T. (2003), "Exploring the Usefulness Kelly's Personal Construct Theory in Assessing Students Learning in Science Courses", *Teaching in Higher Education*, 8/3, 431-445.

Lambert, D. & Lines, D. (2001), *Understanding Assessment: Purposes, Perceptions, Practices,* London, Routledge Falmer.

Lang, J. W. B. & Lang, J. (2010), "Priming Competence Diminishes the Link Between Cognitive Test Anxiety and Test Performance: Implications for the Interpretation of Test Scores", *Psychological Science 21,* 811-819, first published on April 30, 2010.

Lowe, P. A., Lee, S. W., Witteborg, K. M., Prichard, K. W., Luhr, M. E., Cullinan, C. M., et al. (2008), "The Test Anxiety Inventory for Children and Adolescents (TAICA): Examination of the Psychometric Properties of a New Multidimensional Measure of Test Anxiety Among Elementary and Secondary School Students", *Journal of Psychoeducational Assessment*, *26/3*. 215-230.

Lutz, S., & Huitt, W. (2004), "Connecting cognitive development and constructivism: Implications from theory for instruction and assessment. Constructivism in the Human Sciences"*,* 9/1, 67-90.

Lynch, T. & Maclean, J. (2003), *Effects of Feedback on Performance: A Study of Advanced Learners on an ESP Speaking Course*, Edinburg Working Papers in Applied Linguistics.

Marshal, J. M. (2005), *Formative Assessment: Mapping the Road to Success*, Online Document. Retrieved February 24, 2009 from http://www.dcsclients.com/~tprk12/Research_Formative%20Assessment_White_Paper.pdf

Na, Z. (2007). A Study of High School Students' English Learning Anxiety. *Asian EFL Journal. 9/3*. 22-34.

Narciss, S., & Huth, K. (2004), "How to design informative tutoring feedback for multimedia learning", In H. M. Niegemann, D. Leutner, & R. Brunken (Ed.), *Instructional design for multimedia learning,* (pp. 181–195). Munster, New York: Waxmann.

Nicol, D. J. & Macfarlane-Dick, Debra. (2006), "Formative Assessment and self-regulated learning: A model and seven principles of good feedback practice", *Studies in Higher Education*, 31/2, 199-218.

Piaget, J. (2001), *The psychology of intelligence* (2nd Ed.), London: Routledge. [Originally published in 1950]. Retrieved October 12, 2009 from http://www.google.com/books?hl=tr&lr=&id=jJZ7aWY4tIcC&oi=fnd&pg=PR5&dq=The+psychology+of+intelligence&ots=3ZFbpwKbbv&sig=wT5iPnoGw8CeuZiQ17vG1PIQBmA#v=onepage&q=&f=false

Puig, E. (2003), "Assessment and Vygotsky's Zone of Proximal Development", *FLaRE's Online Literacy and Reading Connection,* 3/25.

Putwain, d. W. (2009), "Situated and Contextual Features of Test Anxiety in UK Adolsecent Students", *School of Psychology International,* 30/1, 56-74.

Razazadeh, M. & Tavakoli, M. (2009), "Investigating the Relationship among Test Anxiety, Gender, Academic Achievement and Years of Study: A Case of Iranian EFL University Students", *English Language Teaching,* 2/4, 68-74.

Ryan, A. M. & Patrick, H. (2001), "The classroom social environment and changes in adolescents' motivation and engagement during middle school", *American Educational Research Journal,* 38/2, 437-460.

Reiss, C. R. (n.d), *The importance of pilot studies in the development of large-scale seawater desalination plants*, Retrieved April 7, 2010 from http://www.twdb.state.tx.us/iwt/desal/docs/The%20Future%20of%20Desalination%20in%20Texas%20-%20Volume%202/documents/C5.pdf

Roos, B. & Hamiton, D. (2004), *Towards Constructivist Approach?,* Paper presented at the Annual Conference of the Nordisk Forening for Pedagogisk Forskning (NFPF), Reykjavik, 10-14 March. Retrieved July 3, 2008 from http://www.onlineassessment.nu/onlineas_webb/products/NERA_paper_040217.pdf

Ross, S. J. (2005), "The Impact of Assessment Method on Foreign Language Proficiency Growth", *Applied Linguistics,* 26/3, 317-342.

Sadler, D.R. (1989), "Formative Assessment and The Design of Instructional Systems", *Instructional Science*, *18/2*, 119-144.

Shahi, M. A. J. (2009), "Anxiety in Language Learning", *ELT Weekly, 34*. Retrieved March 31, 2010 from http://eltweekly.com/more/ELT%20Newsletter/anxiety-in-language-learning/

Shepard, L. A. (2000) *The Role of Classroom Assessment in Teaching and Learning,* Center for the Study of Evaluation (CSE Technical Report # 517). University of California, Los Angles, CA.

Shepard, L. A. (2005), "Linking Formative Assessment to Scaffolding", *Educational Leadership, 63/3*. 66-70.

Shute, V. J. (2007), "Focus on Formative Feedback", *Educational Testing Service*, March. Retrieved November, 14, 2009 from http://www.ets.org/Media/Research/pdf/RR-07-11.pdf

Stansbury, K. (2001), "The Role of Formative Assessment in Induction programs", *Reflections,* 4/2, 1-11.

Stiggins, R. J., Griswold, M. M. & Wikelund, K. R. (1989), "Measuring thinking skills through classroom assessment", *Journal of Educational Measurement*, *26*. 233-246. Retrieved November 7, 2009 from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1c/56/b1.pdf

Stiggins, R. J., & Conklin, N. F. (1992), *In teachers' hands: Investigating the practices of classroom assessment,* Albany, NY: SUNY Press.

Stiggins, R. (1992), "*High Quality Classroom Assessment: What Does it Really Mean?", NCME Instructional Topics in Educational Measurement Series,* Module 12, Summer 1992. (Online) at http://www.ncme.org/pubs/items/19.pdf

Stiggins, R. J. (2002), "Assessment Crisis: The Absence of Assessment FOR Learning", *Online article, Kappan Professional Journal*, Retrieved August 12, 2009 from http://electronicportfolios.org/afl/Stiggins-AssessmentCrisis.pdf.

Stiggins, R. (2007), "Conquering the Formative Assessment Frontier", In J. H. McMillan. (2007), *Formative Classroom Assessment-Theory into Practice*, 8-29, Teachers College Press. New York.

Stoneman, B. W. H. (2005), "The Impact of An Exit English Test On Hong Kong Undergraduates: A Study Investigating The Effect Of Test Status On Students' Test Preparation Behaviors", *Unpublished PhD Dissertation,* The Department of English, the Hong Kong Polytechnic University.

Struyven, K., Dochy, F. and Janssen, S. (2005), "Students' Perceptions About Evaluation And Assessment In Higher Education: A Review", *Assessment and Evaluation in Higher Education,* 30/4, 331–347.

Svihla, V. (2006), "Formative assessment: reducing math phobia and related test anxiety in a geology class for non-science majors", *Proceedings of the 7th international conference on learning sciences.*

Taras, M. (2005), "Assessment- Summative and Formative- Some Theoretical Reflections", *British Journal of Educational Studies,* 53/4, 466-478.

Taylor, S. J. & Bogdan, R. (1998), *Introduction to Qualitative Research Methods: a guidebook and resource*, 3rd edition, John Wiley & Sons, Inc. USA.

Teijlingen, E. R. & Hundley, V. (2001), "The Importance of Pilot Studies", *Social Research Update,* 35, Retrieved April 7, 2010 from http://auraserv.abdn.ac.uk:9080/aura/bitstream/2164/157/1/SRU35%2Bpilot%2Bstudies.pdf

Thaine, C. (2004), "The Assessment of Second Language Learning", ELT Journal, 58/4, October, 336-345.

Threlfall, J. (2005), "The Formative Use of Assessment Information in Planning – The Notion of Contingent Planning", *British Journal of Educational Studies*, 53/1, 54–65.

Unruh, S. M. & Lowe, P. A. (2010), "The Development and Validation of a Spanish Language Version of the Test Anxiety Inventory for Children and Adolescents", *Hispanic Journal of Behavioral Sciences,* 32/1, 164-183.

Valette, R. (1994), "Teaching, testing and assessment: Conceptualizing the relationship", In C. Hancock (ed.) *Teaching, testing and assessment: Making the connection*, Lincolnwood, IL: National Textbook Company. 1-42.

Vygotsky, L. S. (1978), *Mind in society: The development of higher psychological processes*, Cambridge, MA: Harvard University Press (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. And Trans.; original works published 1930–1935).

Wachelka, D., Roger, C. K. (1999). Reducing test anxiety and improving academic self-esteem in high school and college students with learning disabilities. *Journal of Behavior Therapy and Experimental Psychiatry. 30*. 191-198.

Wang. T. H. (2007), "What strategies are effective for formative assessment in an e-learning environment?", *Journal of Computer Assisted Learning,* 23, 171–186.

Watering, G., Gijbels, D., Dochy, F., Rijt, J. (2008), "Students' Assessment Preferences, Perceptions of Assessment and Their Relationships to Study Results", *High Education,* 56, 645-658.

Wiliam, D. & Leahy, S. (2007), "A Theoretical Foundation for Formative Assessment", In J. H. McMillan. (2007), *Formative Classroom Assessment: From Theory to Practice*, *29-43*, Teachers College Press, New York.

Windschitl, M. (2002), "Framing Constructivism in Practice as the Negotiation of Dilemmas: An Analysis of the Conceptual, Pedagogical, Cultural, and Political Challenges Facing Teachers", *Review of Educational Research,* 72/2, 131–175.

Wojtczak, A. (2002), "Glossary of Medical Education Terms", Retrieved May 13, 2009 from http://www.iime.org/glossary.htm

Young, E. (2005), *Assessment for Learning: Embedding and Extending*, Retrieved November 5, 2009 from http://www.ltscotland.org.uk/Images/Assessment%20for%20Learning%20version%202vp_tcm4-385008.pdf

Zuriff, G. (1997), "Accommodations for test anxiety under ADA", *Journal of American Academy Psychiatry and the Law online. 25:2.* 197-206. Retrieved September 3, 2009 from http://www.jaapl.org/cgi/content/abstract/25/2/197

APPENDIX 1. TEST ANXIETY INVENTORY

Directions: Read each items below to see if it reflects your experience in test taking. If it does, place a check mark on the line next to the number of the statement. Check as many as seem fitting. Be honest with yourself.

1- Getting a good score on one test does not seem to increase my confidence on other tests _____

2- During a test, I sometimes find myself having trains of thought that have nothing to do with the test_____

3- I do not enjoy eating before or after an important test_____

4- It seems to me that test sessions should not be made the formal, tense situations they are_____

5- People who do well on tests generally end up in better positions in life_____

6- Having to face an important test disturbs my sleep_____

7- If exams could be done away with, I think I would actually learn more from my courses_____

8- Knowing that my future depends in part on doing well on tests upsets me_____

9- People will question my ability if I do poorly_____

10- I never seem to be fully prepared to take tests_____

11- I mentally freeze up on important tests_____

12- I have a hollow, uneasy feeling before taking a test_____

13- Tests make me wonder if I will ever reach my goals_____

14- I sometimes seem to defeat myself (think negative thoughts) whole working on an important test_____

15- If I do not do well on a test, I guess it will mean I am not as smart as I though I was_____

16- My anxiety about tests makes me want to avoid preparing fully, and this just makes me more anxious_____

17- I often find my fingers tapping or my legs jiggling while taking a test_____

18- When taking a test, my emotional feelings interfere with my concentration_____

19- The harder I work on some test items, the more confused I get_____

20- I do not feel confident and mentally relaxed before a test_____

21- My friends will be disappointed in me if my score is low_____

22- One of my problems is not knowing exactly when I am prepared for a test___

23- I often fell physically panicky when I have to take a really important test_____

24- I am going to find out how others did before I announce my score_____

25- Some people I know will be amused if I score low, and this bothers me_____

26- My test performance is directly connected to my future success and security_____

27- During tests, I sometimes get so nervous that I forget facts I really know_____

APPENDIX 2. TEST ANXIETY INVENTORY (ORIGINAL)

Directions: Read each items below to see if it reflects your experience in test taking. If it does, place a check mark on the line next to the number of the statement. Check as many as seem fitting. Be honest with yourself.

___ 1. I wish there were some way to succeed without taking tests.

___ 2. Getting a good score on one test does not seem to increase my confidence on other tests.

___ 3. People (family, friends, etc.) are counting on me to do well.

___ 4. During a test, I sometimes find myself having trains of thought that have nothing to do with the test.

___ 5. I do not enjoy eating before or after an important test.

___ 6. I have always dreaded courses in which the teacher has the habit of giving pop quizzes.

___ 7. It seems to me that test sessions should not be made the formal, tense situations they are.

___ 8. People who do well on tests generally end up in better positions in life.

___ 9. Before or during an important exam, I find myself thinking about how much brighter some of the other test-takers are.

___ 10. Event though I don't always think about it, I am concerned about how others will view me if I do poorly.

___ 11. Worrying about how well I will do interferes with my preparation and performance on tests.

___ 12. Having to face an important test disturbs my sleep.

___ 13. I cannot stand to have people walking around watching me while I take a test.

___ 14. If exams could be done away with, I think I would actually learn more from my courses.

___ 15. Knowing that my future depends in part on doing well on tests upsets me.

___ 16. I know I could outscore most people if I could just get myself together.

___ 17. People will question my ability if I do poorly.

___ 18. I never seem to be fully prepared to take tests.

___ 19. I cannot relax physically before a test.

___ 20. I mentally freeze up on important tests.

___ 21. Room noises (from lights, heating/cooling systems, other test-takers) bother me.

___ 22. I have a hollow, uneasy feeling before taking a test.

___ 23. Tests make me wonder if I will ever reach my goals.

___ 24. Tests do not really show how much a person knows.

___ 25. If I score low, I am not going to tell anyone exactly what my score was.

___ 26. I often fell the need to cram before a test.

___ 27. My stomach becomes upset before important tests.

___ 28. I sometimes seem to defeat myself (think negative thoughts) whole working on an important test.

___ 29. I start feeling very anxious or uneasy just before getting test results.

___ 30. I wish I could get into a vocation that does not require tests for entrance.

___ 31. If I do not do well on a test, I guess it will mean I am not as smart as I though I was.

___ 32. If my score is low, my parents will be very disappointed.

___ 33. My anxiety about tests makes me want to avoid preparing fully, and this just makes me more anxious.

___ 34. I often find my fingers tapping or my legs jiggling while taking a test.

___ 35. After taking a test, I often feel I could have done better than I actually did.

___ 36. When taking a test, my emotional feelings interfere with my concentration.

___ 37. The harder I work on some test items, the more confused I get.

___ 38. Aside from what others may think of me, I am concerned about my own opinion of myself if I do poorly.

___ 39. My muscles tense up in certain areas of my body when I take a test.

___ 40. I do not feel confident and mentally relaxed before a test.

___ 41. My friends will be disappointed in me if my score is low.

___ 42. One of my problems is not knowing exactly when I am prepared for a test.

___ 43. I often fell physically panicky when I have to take a really important test.

___ 44. I wish teachers understood that some people are more nervous than others when taking tests, and that this could be taken into account when test answers are evaluated.

___ 45. I would rather write a paper than take a test for a grade.

___ 46. I am going to find out how others did before I announce my score.

___ 47. Some people I know will be amused if I score low, and this bothers me.

___ 48. I think I could do much better on tests if I could take them alone and/or not feel pressured by a time limit.

___ 49. My test performance is directly connected to my future success and security.

___ 50. During tests, I sometimes get so nervous that I forget facts I really know.

Test Anxiety Inventory Analysis

Four Main Sources of Test Anxiety:

1. Concerns about how others will view you if you do poorly

· Items 3, 10, 17, 25, 32, 41, 46, 47

2. Concerns about your own self-image

· Items 2, 9, 16, 24, 31, 38, 40

3. Concerns about your future security

· Items 1, 8, 15, 23, 30, 49

4. Concerns about not being prepared for a test

· Items 6, 11, 18, 26, 33, 42

Three Main Expressions of Test Anxiety:

1. Bodily reactions

· Items 5, 12, 19, 27, 34, 39, 43

2. Thought disruptions

· Items 4, 13, 20, 21, 28, 35, 36, 37, 48, 50

3. General test-taking anxiety

· Items 7, 14, 22, 29, 44, 45

APPENDIX 3. ASSESSMENT PREFERENCE SCALE

- **Look at the assessment types below used for evaluating students' course studies, and choose the ones that you would prefer your instructors to use for your assessment procedure.**

5	**4**	**3**	**2**	**1**
ALWAYS	OFTEN	SOMETIMES	RARELY	NEVER

a) Matching ……
b) Multiple choices …….
c) Open-ended ……
d) Rewrite ……
e) Clause test ……
f) Explanatory questions ……..
g) Combining sentences …….
h) Self assessment ……
i) Peer assessment ……..
j) Receiving feed back from teacher ……
1- Written feedback …….
2- Verbal feedback ……

k) Ongoing assessment ……
1- Weekly quiz …….
2- Classroom activities ……

l) Take-home tasks …….

APPENDIX 4. CHECKLIST FOR STUDENTS

I CAN IDENTIFY SENTENCES THAT CONSIST OF:	**YES**	**PARTLY**	**I HAVE NO IDEA**
The Structure of Sentenses			
Two-Part Sentences			
Sentence Patterns			
"Sentence" and "Clause"			
Cohesion			
Known and New Contract			
Metadiscourse			
Parallelism			
Levels of Generality			
Sentence Rhythm			
Intonation			
Controlling Rhythm			
Writer's Voice			
Contradictions			
Verbs and Formality			

APPENDIX 5. A SAMPLE FOR WEEKLY QUIZ

Select the option that best describes the arrangement of clauses within each sentence. Sometimes the description will refer to only the underlined part of a sentence. Read all the options carefully before making your choice.

1. Before Raoul bought his new car, he remembered that his grandmother might give him her old sedan.

- The independent clause is "Before Raoul bought his new car."
- The independent clause in this sentence contains only two words.
- There are two independent clauses here.

2. Alicia's wrist began to hurt, and her mother made a doctor's appointment for her.

- The underlined part is the sole independent clause here.
- The real independent clause begins with "and."
- This sentence contains two independent clauses.

3. Zipping around the park in her little sportster, Juditha began to feel, at last, like a millionaire.

- The underlined part is the independent clause.
- Trick question: there is no independent clause.
- The independent clause begins with "Juditha."

4. Although Jerzy had taken several courses in computer science, he couldn't solve the problems with his hard drive.

- The underlined part is the independent clause.
- The independent clause begins with the word "he."
- The sentence contains two independent clauses.

5. The Celtics have begun their long journey back to championship basketball, working seriously on fundamentals and beginning to attract new fans to the game.

- The sentence is one long independent clause.
- The sentence contains two independent clauses.
- Only the underlined part is the independent clause.

6. Ernesto wanted to spend the night studying, but his wife had other ideas.

- ☐ The sentence contains two independent clauses.
- ☐ Only the underlined part of the sentence is an independent clause.
- ☐ The sentence is one long independent clause (with a compound verb).

7. The crowd began to clap and cheer as the team entered through a cloud of smoke.

- ☐ The sentence contains two independent clauses.
- ☐ The independent clause begins with the word "as."
- ☐ The underlined part is the independent clause.

8. Gordy worked his way up to middle management but kept on wishing for a better job.

- ☐ The underlined part of the sentence is the independent clause.
- ☐ The sentence is one independent clause.
- ☐ The sentence contains two independent clauses.

9. Gesualdo began to leap around the room when his wife fed him the jalapeño peppers.

- ☐ The independent clause is the part of the sentence not underlined.
- ☐ The independent clause is the part of the sentence that is underlined.
- ☐ The sentence contains two independent clauses

10. Heitor signed the contract but never agreed to purchase the CDs.

- ☐ This sentence contains two independent clauses.
- ☐ The underlined part is the complete independent clause.
- ☐ The entire sentence is one independent clause.

CURRICULUM VITAE

PERSONAL CONTACT

Name : Kağan BÜYÜKKARCI
E-mail : kagan_fb@hotmail.com
Marital Status : Married
Date of Birth : 16.02.1982

EDUCATION AKADEMIC

2006-2010 : Cukurova University PhD, Institute of Educational Sciences English Language Teaching, Adana.
2004 - 2006 : Selcuk University MA, Institute of Educational Sciences English Language Teaching, Konya.
2002-2003 : University of Houston BA
2000 - 2004 : Selcuk University BA**,** Selcuk Faculty of Education English Language Teaching, Konya.

EXPERIENCE

2006/2010 : Cukurova University, ELT Department Research Assistant, Adana
2004/2006 : Selcuk University, ELT Department Research Assistant, Konya

PUBLISHED ARTICLES

Genc, B. & Buyukkarci, K. (2009). Discourse Markers: The Case of *AND* in the Speech of Turkish Speakers of English. *Linguistics Journal 4(3)*.

Buyukkarci, K. (2009). A Critical Analysis of Task-Based Learning. *Kastamonu Education Journal 17 (1)*, 313-320.

Baykul, Y., Gürsel, M., Sulak, H., Ertekin, E., Yazıcı, E., Dülger, O., Aslan, Y., Büyükkarcı, K. (2010). A Validity and Reliability Study of Grash-Reichman Student Learning Style Scale. *International Journal of Social Sciences. 5 (3)*. 177-184.

Buyukkarci, K. (2010). Teaching Phrasal Verbs Through Communicative Approach. *Kafkas University Journal of the Institute of Social Sciences.*

Genç, B., Buyukkarci, K. (In Press). The Usage of Deixis in the Speeches of Turkish Speakers of English. *Kafkas Eğitim Fakültesi Dergisi*.

CONGRESSES and PROGRAMS ATTENDED AS A SPEAKER

5th Annual EALTA Conference-Hellenic American Union-"Assessment Literacy in Europe and Beyond: Realities and Prospects", ***(A Case Study: Conceptual Changes in Assessment Practices)*** **Athens-2008.**

Cukurova University Faculty of Education- "Going Beyond Methods: Active Teaching Strategies" ***(Formative Assessment)*** **Adana- 2008.**

The 11th International INGED ELT Conference at Ankara University-"Stretching the Boundaries", (***Student-Teacher Attitudes towards Students' Behaviors)*** **Ankara-2007.**

CONGRESSES, WORKSHOPS AND PROGRAMS ATTENDED

* The 6th Annual EALTA Conference-"Synergies and Tensions in Language Testing and Assessment" University of Turku, Turku-2009
* 6th EALTA Pre-conference Workshop-"Introduction to Item Response Theory" Turku-2009
* 5th EALTA Pre-conference Workshop-"Classroom Assessment" Athens-2008
* Cukurova University Life Long Learning, Intensive Program-"Intercultural Dialogue" Adana-2008
* Cukurova University Center for Foreign Languages-"Stepping Over Thresholds: Transitions in English Language Learning and Teaching" Adana-2008
* Mersin University Foreign Languages Center-"The Use of Course Books in Language Teaching" Mersin-2007
* ICANAS 38 "International Congress of Asian North African Studies" Ankara-2007
* The Swedish Handicap Institute-"Children's Train Exhibition" Adana-2007
* 3rd National Meeting of Vocational High School Administrators- Adana-2007
* 3rd Social Sciences Teaching Congress - Adana-2007

* INGED Event-"Different Activities for Different Age Groups" Ankara-2007
* Cukurova University English Language Teaching Depaertment-"4th International Postgraduate Conference in Linguistics and Language Teaching" Adana-2006

MEMBERSHIPS TO ACADEMIC ASSOCIATIONS

2007 -... : EALTA (European Association for Language Testing and Assessment) Affiliate Member

2006 – 2007 : INGED (Turkey's ELT Association) Member