T. C.
PAMUKKALE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
YABANCI DİLLER EĞİTİMİ ANABİLİM DALI
İNGİLİZ DİLİ EĞİTİMİ BİLİM DALI
YÜKSEK LİSANS TEZİ

AN INVESTIGATION INTO TESTING AND ASSESSMENT

IN YOUNG LEARNERS' CLASSROOMS: DOES PRACTICE

MATCH THE POLICY?

Meral ÜÇOK ATASOY

Denizli – 2019

T.R.
PAMUKKALE UNIVERSITY
INSTITUTE OF EDUCATIONAL SCIENCES
DEPARTMENT OF FOREIGN LANGUAGE EDUCATION
ENGLISH LANGUAGE TEACHING PROGRAM
MASTER OF ARTS THESIS

AN INVESTIGATION INTO TESTING AND ASSESSMENT IN YOUNG LEARNERS' CLASSROOMS: DOES PRACTICE MATCH THE POLICY?

Meral ÜÇOK ATASOY

Supervisor

Assoc. Prof. Dr. Recep Şahin ARSLAN

JÜRİ ÜYELERİ ONAY SAYFASI

Bu çalışma, İngiliz Dili Eğitimi Anabilim Dalı'nda jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Başkan:. Prof. Dr. Şevki Kömür

Üye: Prof. Dr. Turan PAKER

Üye: Doç. Dr. Recep Şahin ARSLAN (Danışman)

Pamukkale Üniversitesi Eğitim Bilimleri Enstitüsü Yönetim Kurulu'nun .17.07.2019 tarih ve .30.1.9....sayılı kararı ile onaylanmıştır.
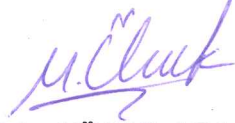
Prof. Dr. Mustafa BULUŞ

Enstitü Müdürü

ETİK BEYANNAMESİ

Pamukkale Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

> Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,

> Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,

> Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,

> Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,

> Kullanılan verilerde herhangi bir tahrifat yapmadığımı,

> Bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı beyan ederim.

Meral ÜÇOK ATASOY

ACKNOWLEDGEMENTS

DEDICATION

To my parents and

    my beloved ones Mehmet, İldeniz and Bilge

ÖZET

Küçük Yaşta Yabancı Dil Öğrenen Öğrencilere Uygulanan Ölçme ve Değerlendirme Üzerine Bir İnceleme: Uygulamalar Yabancı Dil Öğretim Politikasına Uygun mu?

ÜÇOK ATASOY, Meral

Yüksek Lisans Tezi, Yabancı Diller Eğitimi Anabilim Dalı,
İngiliz Dili Eğitimi Bilim Dalı
Danışman: Doç. Dr. Recep Şahin ARSLAN
Haziran 2019, 101 sayfa

Son yıllarda ölçme ve değerlendirme, gerek araştırmacılar gerekse yabancı dil öğretiminde söz sahibi olan kişi ve kurumlar tarafından giderek daha çok ilgi görmeye başlamıştır. Ölçme ve değerlendirme, yabancı dil öğretme ve öğrenme uygulamaları için bir ayna niteliği taşıdığı ve etkililikleri konusunda geri dönüt sağladığı için yabancı dil öğretiminde ikinci planda bırakılmamalıdır. İngilizce öğrenme sürecinin başlangıcında olan küçük yaştaki öğrenciler, doğaları gereği yetişkinlerden farklıdırlar. Bu nedenle bu öğrencilere uygulanan ölçme ve değerlendirme uygulamaları büyük dikkat gerektirmektedir. Bu nedenle, dil öğretimi konusunda hem ulusal hem de uluslararası alanda sürekli yenilikler yapılmaktadır. Dışarıdan bakıldığında her şey yolunda görülse de, sınıf içi uygulamalarda durumun olması gerekene uygun olup olmadığının anlaşılması gereklidir. Bu bilinçle, bu çalışma ortaokullarda eğitim veren İngilizce öğretmenlerinin ölçme ve değerlendirme uygulamaları üzerine bir inceleme niteliği taşımaktadır. Bu çalışma ayrıca Milli Eğitim Bakanlığı tarafından oluşturulan İngilizce Öğretim Programı ile ortaokullarda eğitim veren İngilizce öğretmenlerinin ölçme değerlendirme uygulamaları arasındaki tutarlılığı incelemektedir. Bu çalışma 2017-2018 eğitim-öğretim yılı bahar dönemi sonunda gerçekleştirilmiştir. Bu araştırma Denizli'nin merkez ilçelerindeki ortaokullarda eğitim veren 152 İngilizce öğretmeninin katılımıyla gerçekleştirilmiştir. Araştırmanın verileri nicel ve nitel olmak üzere beşli Likert tipi anketler ve öğretmenlerin İngilizce ölçme değerlendirmelerinde kullandıkları sınav kâğıtları yoluyla toplanmıştır. Anketler üzerindeki verilerin incelenmesinde SPSS 24 betimleyici istatistik analizi, sınav kâğıtlarının incelenmesinde içerik ve doküman analizi uygulanmıştır. Bulgular İngilizce

Öğretim Programıyla İngilizce öğretmenlerinin ölçme değerlendirme uygulamaları arasında tutarsızlık olduğunu göstermiştir. İngilizce öğretmenlerinin iletişimsel yeterlilikten çok dil bilgisi ağırlıklı klasik sınavlar hazırladıkları saptanmıştır.

Anahtar kelimeler: Ölçme ve değerlendirme, yabancı dil olarak İngilizce, küçük yaştaki öğrenciler, ortaokullar, eğitim politikası, uygulamalar.

ABSTRACT

An Investigation into Testing and Assessment in Young Learners' Classrooms: Does Practice Match the Policy?

ÜÇOK ATASOY, Meral

Master Thesis, Department of Foreign Languages Education,
English Language Teaching Program
Supervisor: Assoc. Prof. Dr. Recep Şahin ARSLAN
June 2019, 101 pages

In the last decades, testing and assessment have gained increasing attention by both researchers and the stakeholders in the field of language teaching. On the grounds that assessment stands for a mirror of teaching and learning practices and feedback about their effectiveness, it could never be ignored or be of secondary importance in language teaching and learning. Another issue of rising attention in the field is young learners. That group of language learners who are at the preliminary stages differ from adult learners in nature. Thus, the practices of teaching and assessment in young learners' classrooms require great care. Based on this, there exist continuous revolutions in the language teaching policies both nationally and internationally. When looking at the situation it seems promising. However, it is significant to look inside the classrooms to realize whether the actual performance reflects the ideal one. With this awareness, this paper represents an investigation of EFL teachers' practices of assessment in young learners' classrooms. Furthermore, it attempts to find out the consistency between the policy and EFL teachers' in-class practices of assessment in lower-secondary schools. This study was conducted at the end of the spring term of 2017-2018 academic year. The participants were 152 EFL teachers working in lower-secondary schools in the central districts of Denizli. Data were collected via five-point Likert-scale questionnaires and teachers' assessment documents. SPSS 24 was applied for the descriptive analysis of data on the questionnaires; document and content analysis were applied for the data on the assessment documents. Results indicated inconsistency between the policy and practices of EFL assessment in lower-secondary schools. EFL teachers tended to design traditional paper and pencil tests based on language structure rather than testing their communicative competence.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

This chapter provides information about the reasons for conducting this study by stating the problem, purpose, significance, and limitations of the study in addition to the research questions and assumptions in line with the previous studies in the field.

Background to the Study

Being the vital element of communication, language is the way to keep connection with the world in all senses. The booming technology and the people's need to catch up with the unignorable consequences of it have contributed to the importance of language in recent decades. As a result, mastering mother tongue- or using it properly in all contexts is not sufficient even in the native countries of people. The reason is that in all lines of business, economics, politics and education a second or foreign language is considered necessary. Other than that, the situation is more challenging in the international scene: a 'second language' and foreign language is required in most cases. At the present time, the language which serves that purpose is English having fame as 'lingua franca' (Crystal, 1997; Harmer, 2007). Consequently, countries including Turkey give great importance to English Language Teaching (ELT). With the aim of improving the English language level, innovations and revisions in the educational policies of countries are continuous. On the other hand, teachers and students, having the major part in the field, have the responsibility to implement teaching and learning in consistency with the policy. This study seeks to find out the current situation in ELT programs in Turkish context examining its extent of match with the classroom practices of teachers in terms of testing and assessment.

In Turkey, Ministry of National Education (MoNE) is responsible for the supervision of public and private education under a national curriculum. MoNE (2018) provides an ELT program for all levels of compulsory education and the language education policy of MoNE is based on the Common European Framework of References for Languages (CEFR) (2001, 2018), which is an international standard for describing language ability (MoNE, 2013, 2015, 2018). This study makes reference to CEFR in order to clarify MoNE's suggested practices of testing and assessment in English as a Foreign Language (EFL) teaching at the lower-secondary levels.

Despite the fact that government policies and curricula typically plan teaching communicatively, this approach most often forms a contrast to the requirements of national structural examinations. This situation can lead to the negative backwash effect as teachers

are under pressure to complete the syllabus in a limited time and prepare students for examinations (Carless, 2003). As McKay (2006) points out, teachers may create an appealing atmosphere and inspire students to be engaged in language keenly; still assessment could ruin all. Namely, even if the teachers are qualified enough and use communicative practices in their classes effectively, inappropriate assessment practices may reverse the situation.

In language education, in addition to the policy's role as a determiner in testing and assessment practices, another determiner is the practices of language teaching and learning. Teaching and learning practices and testing/assessment practices go hand-in-hand in all kinds of education and language education as well. According to Hughes (2003), language assessment and teaching program should be consistent with each other in terms of learning objectives, the kinds of tasks which the children are expected to perform and the content of teaching. By this way, assessment is not something separate from learning. Alderson and Wall (1993) state that tests are powerful determiners to see what happens in the classroom.

 It is important for teachers to understand the reasons and theoretical considerations behind these changes. However, it is not enough to understand the policy as it should be, also how teachers implement them is important. In this regard, Fullan (1993) highlights the value of teachers' role in the changes and innovations of educational programs since they have the responsibility to transfer these changes to the classrooms. In the light of this information it can be stated that language assessment techniques and tools preferred by a language teacher are assumed to mirror his/her teaching practices as well as perceptions about language teaching and learning. In this study, the main focus is on the testing and assessment practices of EFL teachers working in state lower-secondary schools, however, some inferences are made about the teaching practices and perceptions of target EFL teachers.

1.2. Statement of the Problem

In education, no matter what the teaching field is, the objectives of the educational program and the course content should be well fitted to each other. Otherwise, there arise huge gaps between the policy and the implementations of it. Additionally, it becomes impossible to talk about uniformity among the institutions- consequently the teachers in terms of teaching.

In language teachers' preferences of language teaching and testing practices, the educational policy and the language curriculum play a constitutive role. In addition to

recommended techniques in the foreign language curriculum, the suggested materials may be effective in teachers' preferences of teaching and testing techniques. In that matter, it is significant that EFL teachers should follow the innovations in Foreign Language Education (FLE) and adopt their teaching and testing techniques accordingly. In addition, there should be consistency between policy and practice in both language teaching and language testing.

In young learners' classrooms- as the ELT program of MoNE suggests- there should be an authentic atmosphere. Students should be the explorers of the knowledge instead of passive receptors of it. The tasks and materials should promote their communicative competence. In other words, there should emerge a need for children to use the language. Furthermore, while the children dealing with the presented tasks and activities, they should enjoy, laugh, move, jump and see their own as well as one another's progress in developing the language skills. Namely, the classrooms full of young language learners (YLLs) should be alive and kicking. In order to create such an atmosphere, teachers should utilize games, songs, Total Physical Response (TPR) activities, role-plays, presentations, authentic materials and interactive technologies.

In such classrooms where the communicative competence is the main objective of the course, assessment types and techniques should be in accordance with these objectives and the content of the course. As Tsagari (2004) argues, both teaching and assessment should be organized so as to involve students in cognitive skills of analysis and synthesis as higher-order thinking skills. In addition to effective integration of four skills, alternative assessment tools such as self and peer assessment, projects, products, portfolios, role-plays and so should be utilized as well as the traditional and formal tools in order to promote students' communicative competence.

Unfortunately there is a real problem of ELT in Turkey especially in primary and secondary schools: Teachers are willing to apply several teaching methods, tasks, activities based on communicative language teaching while teaching English to young learners and they include four skills in addition to the other linguistic components such as grammar and vocabulary in their lessons; however, when it comes to compose tests in line with their teaching, it is observed that some problems arise and teachers may tend to use more traditional tests based on mostly grammar instead of utilizing several types of assessment tools involving the assessment of four skills. One of the possible reasons is the High School Placement Test (LGS) – a standardized test offered by MoNE which involves only multiple choice items in contrast with its own objectives of ELT curriculum. This

inconsistency between MoNE's own objectives and implementations of testing and assessment is a matter of another research. Another possible reason for teachers' inconsistent assessment practices is the limited time allocated to ELT courses in the curriculum. Teachers have trouble in including four skills in the tests they apply within a semester. Regardless of what the underlying reasons are, the mismatch between the policy and the practices of assessment brings out negative backwash effect to students' language learning and undesirable consequences.

As Hughes (2003) argues, it is not fair and ethical to base the assessment on different content and objectives from the suggested communicative approach of the ELT program. Based on this point of view, this study seeks to find out whether there is a consistency or not between the practices of EFL teachers of young learners and the CEFR-oriented ELT Program of MoNE in terms of testing and assessment practices. Aiming to find some answers to this question, the participants of the study were selected from the EFL teachers working in state lower-secondary schools.

1.3. Purpose of the Study

One of the purposes of this study is to find out what kind of testing and assessment practices EFL teachers apply. Another purpose of the study is to find out whether EFL teachers assess four skills of young EFL learners or not. Main purpose of this study is to find out the extent of consistency between the ELT Curriculum for Primary Schools (for 2nd, 3rd, 4th, 5th, 6th, 7th, and 8th Grades) proposed by MoNE and testing and assessment practices of EFL teachers working in state lower-secondary schools (5th, 6th, 7th, and 8th Grades).

1.4. Research Questions

This study seeks to answer to the following questions:

1. What are the testing and assessment practices of EFL teachers working in state lower-secondary schools?

1.a How frequently do the EFL teachers prefer traditional paper-pencil tests and alternative ways of assessment?

1.b Are there any differences among teachers' preferences of testing and assessment tools in terms of their demographic features of gender, experience and the highest degree they hold?

1.c Which language skills of young EFL learners are assessed by EFL teachers at state lower-secondary schools?

2. To what extent are the testing and assessment practices of EFL teachers consistent with the course outcomes stated by the Ministry of National Education in the English Language Teaching Program for the 5th, 6th, 7th and 8th grades?

1.5. Significance of the Study

Language education is one of the indispensable components of general education all over the world. Without learning a second or foreign language it is almost impossible for one to get a good job, have a comfortable life, be a respected person in the society, etc. Teaching English as a second or foreign language to young learners is a significant matter in today's world on the ground that the children are the guarantee of the future. In order to be sure of the quality of language education that they are exposed to, the policy and the teachers as the implementers of the education should work in consistency. Being the core elements of teaching and learning processes, testing and assessment are the keys to see the effectiveness of the teaching, the materials and tasks, and finally the goals of the ELT program. To be able to recognize the extent of consistency between the assessment practices suggested in the curriculum and implemented in the young language learners (YLLs) classroom, it is important to conduct studies examining the real conditions at schools. In that sense, this study sheds light on EFL teachers' assessment practices at state lower-secondary schools examining them in terms of fitting in with communicative competence and communicative language testing as suggested in the ELT program.

 If there is a mismatch between teachers' practices of testing and assessment and the ELT Program, the reasons behind it could be investigated thoroughly, and solutions could be produced accordingly. In that sense, this study gives ways to further research. The inconsistency may stem from the ELT teachers' lack of pedagogical content knowledge the Ministry's high stake exams, the limitation of class-period of English lessons or supplied materials inter alia. As taking these reasons into account, the results of this study could be used to make some changes in the ELT program or the examination system of the Ministry; teachers may be provided with in-service training, or the allocated time for EFL teaching may be broadened.

Since the new policy has been implemented in Turkey since 2012, for only seven years, the number of studies on teachers' practices in both teaching and testing English in

young learners' classrooms is limited to a small number in Turkish context. This study will contribute to research on language testing and assessment and EFL teaching in Turkey. It may also help language teachers, teacher trainers, and curriculum developers better understand the current situation from teachers' perspectives in terms of practices of foreign language testing in young learners' classrooms. Over and above, this study constitutes a model for future researchers to be inspired and encouraged to conduct studies on the actual EFL teaching and learning settings both in national and international contexts. In the long run, this type of studies can be valuable in helping EFL instruction to better meet the needs of the growing population of young EFL learners.

1.6. Limitations of the Study

This study is conducted with the aim of investigating the testing and assessment practices of EFL teachers working with young learners at state lower-secondary schools. The following limitations can be listed for this study:

1. The findings of the study are limited to the randomly selected EFL teachers working in state lower-secondary schools during the 2017-2018 academic year.

2. The number of the participants may not be sufficient to reflect the whole picture of teachers' practices of EFL testing and assessment in Turkey.

1.7. Assumptions of the Study

Main assumptions of the study are listed as follows:

1. It is assumed that all the participants participated in the study willingly and responded sincerely to the data collection instrument.
2. The number of the participants is adequate to represent all EFL teachers working in secondary schools in the central districts of Denizli.
3. The findings of the study would present the real condition in testing and assessment practices of EFL teachers working in lower-secondary schools.

CHAPTER II: LITERATURE REVIEW

2.1. Key Concepts in English Language Assessment

2.1.1. Testing and Assessment

Language testing and assessment is one of the prevalent issues of today's educational and scientific era. By providing the necessary feedback and positive backwash effect to improve teaching and learning for teachers or test-makers in addition to the students or the test-takers, assessment practices play an essential role in language teaching and learning as well (Brown, 2007; Bachman and Palmer, 2010; Cheng and Fox, 2017).

In most cases the terms "testing" and "assessment" are used synonymously or interchangeably: gathering information (Bachman and Palmer, 2010). Brown (2007, p.445) makes a more clear distinction: "A test is a method of measuring a person's ability or knowledge in a given domain. Assessment, on the other hand, is an ongoing process that encompasses a much wider domain". A test is a way of measuring students' performance but the results indicate the students' abilities, in other words *competence* (Brown, 2004). In this manner, testing constitutes one of the forms of assessment without being the only one. In order to assess a student's language performance, other procedures and tasks are needed in addition to the tests (Brown, 2007). According to Hughes (2003), teachers need to make meaningful comparisons, so they need to apply tests rather than assessment. In the light of this information, these two terms are used interchangeably throughout this study. Brown (2004, p.5) presents a diagram explaining the relationship between tests, assessment and teaching as can be seen in Figure 2.1.
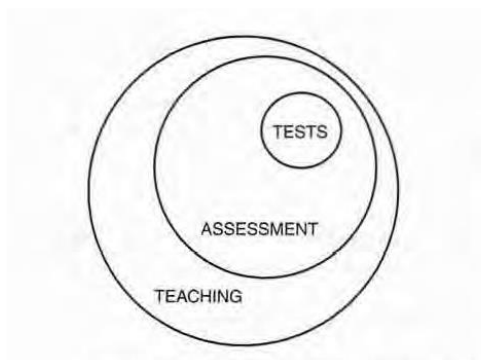
Figure 2.1. Tests, Assessment, and Teaching

2.1.2. Formal vs. Informal Assessment

Formal assessments are systematic, standardized, pre-planned tests aiming to assess students' success at specific content. They can be utilized to compare the results according to particular standards (Brown, 2004). Informal assessment, on the other hand, includes unplanned, spontaneous types of assessment which do not require any recordings and constant decisions. Even the teachers' comment after a completion of a task by students is a form of informal assessment (Brown, 2004). Formal assessment types could be listed as criterion-referenced tests, norm-referenced tests, achievement tests, and aptitude tests.

2.1.3. Formative vs. Summative Assessment

The distinction between formative and summative assessment may be made in terms of purpose and use of information gathered by the assessment. When the assessment provides immediate feedback for ongoing teaching and learning, this type of assessment is formative (Cameron, 2005). This is the classical definition for formative assessment. On the other hand it is known as *assessment for learning* in literature (Black, Harrison, Lee, Marshall & Wiliam, 2003); Hughes (2003) stands out that informal quizzes and tests may serve as a piece of formative assessment as well as the observations or portfolios; even so a harmony of information obtained from different sources should reflect the same result for an individual student.

Bachman and Palmer (2010) state that with the help of feedback by formative assessment, both teachers and students may make changes in their teaching and learning. Teachers may change their way of instruction to send the message better; the students may make decisions about better language learning. For this reason, the purpose of lying behind the formative assessment involves more high-stake decisions instead of low-stake decisions (Mckay, 2006).

On the other hand summative assessment takes place at the end of a unit, a term, a school year or any type of study period. It may be based on the teacher's summative observations of the students or the results of tests formalizing their achievement and focusing on the mastery of linguistic accuracy (Brown, 2004; Shaaban, 2005; Bachman and Palmer; 2010, Mckay, 2006). Summative assessment is limited in making decisions to improve teaching and learning practices since it is applied after the target process of teaching and learning is complete. Nevertheless, they are the common parts of assessment in most of the educational programs as well as MoNE's ELT Program.

Shaaban (2005) argues that in most cases at secondary schools, summative assessment- emphasizing the linguistic competence rather than communicative competence- is in the foreground compared to formative type of assessment. On the other hand, Cheng and Fox (2017) point out that in real classroom environments, teachers apply both *assessment for learning* which is of formative assessment and *assessment of learning* which is of summative assessment; in fact in language education both of them are significant and bring on different advantages for both teachers and students. In line with the literature it would be advisable that without overusing one or another, the best is to utilize both of them regarding students' characteristics and for their progress in both communicative and linguistic competence.

2.1.4. Diagnostic vs. Achievement Tests

In discrimination of these two types of assessment, the purpose of the assessment serves as the determiner. If the purpose of the assessment is to detect what the students are capable of, the type of test is *achievement test*. On the other hand, if the purpose of the assessment is to gather information about students' both capabilities and inabilities individually, and enhance the learning in the long run, this time it is a *diagnostic test* (Cameron, 2005; Brown, 2007; Fulcher and Davidson, 2007). In this manner, it can be stated that diagnostic tests are and should be more preferable than the achievement tests for young learners in order to foster their learning continuously. However, achievement tests are mostly adopted at state schools and divided into two types in itself: *progress achievement tests*, and *final achievement tests*. For progress achievement tests midterm exams conducted mostly two or three times during a semester can be good examples. Final exams conducted at the end of the semester can be a final achievement sample.

2.1.5. Criterion-Referenced vs. Norm-Referenced Tests

If the aim of a test is to describe what a student knows and is able to, it is a criterion-referenced test. This type of test compares the student's achievement to a certain criteria of learning objectives. On the other hand, when the aim of a test is to compare a students with the others, it is a norm-referenced test (Brown, 2004; Cameron, 2005; Cheng and Fox, 2017). Brown (2004) states that the classroom tests which are conducted in a class based on a curriculum are the typical examples of criterion-referenced tests. In addition, Brown (2004) points out that Test of English as a Foreign Language (TOEFL) may be used as a norm-referenced test.

2.2. Basic Principles for Effective Tests

2.2.1. Practicality

"A good test is practical. It is within the means of financial limitations, time constrains, ease of administration, scoring and interpretation" (Brown, 2007, p.446). Practicality is something necessary for testing and assessment within the large amount of linguistic components in teaching and learning process. That practicality is on behalf of both the teachers as test-makers and the students as the test-takers. The practicality of a test depends on the purpose it serves. In that sense, it is important for test-makers to decide whether the test is norm-referenced or criterion-referenced. In norm-referenced tests the test-taker is placed in a numerical order within the other test-takers. In that case, computers do the work for practicality. In the case of criterion-referenced tests, teachers spend much effort to supply the students with the necessary feedback (Brown, 2007).

2.2.2. Reliability

In order to label a test as reliable, it needs to be consistent and dependable (Brown, 2007; Cheng and Fox, 2017). In order for the results to be reliable, the test itself, the student as the test-taker, the administration and the scoring of the test should be exempt from the reliability problems. In classroom-based tests, if the student's mood is in the way to negatively affect the scores, this is beyond the control of the test-maker and it means *student-related unreliability*. Similarly, if the place where the test is administered has some negative effects on the scores, this time is *administration unreliability*. If the test scores are not consistent when scored by more than one scorer, this time it is *scorer unreliability*.

2.2.2.1. Rater-reliability. According to Fulcher (2010), rater-reliability is concurrency among the people who rate a test. Since it requires human-rating Bachman and Palmer (2010) claim that at least two raters are needed to be able to ensure the rater-reliability. In the school context, the raters are the teachers. They are in a continuous effort to contribute learner's improvement. Hence, it is natural that they may feel exhausted and it may decrease the reliability of their ratings. To prevent this, the number of the raters should be increased (Underhill, 1987).

2.2.3. Validity

Validity is the concept which determines the test's suitability to what is actually intended to assess (Henning, 1987). To give an example, if a test is constructed with the

aim of assessing the speaking skills of the learners and it achieves this aim, the test is labelled as 'valid' and vice versa. On the other hand, there is relationship between validity and reliability: reliability is criterion for validity in the sense that (Alderson, Clapham and Wall, 1995; Bachman and Palmer, 2010). Namely, a test can be reliable even if it is not valid. However, it cannot be valid if it is not reliable.

2.2.2.2. Content validity. Content validity ensures that a test covers all the relevant structures of the language and skills (Hughes, 2003). According to Henning (1987), content validity is all about the test's comprehensiveness in terms of the target language components. Hughes (2003) claims that content validity of a test should be ensured in the process of its development; otherwise, it is not possible to ensure content-validity while it is being administered.

2.2.2.3. Criterion-related validity. According to Hughes (2003, p. 27) criterion validity "refers to the degree to which results on the test agree with those provided by some independent and highly dependable assessment of the candidate's ability". The independent test stands for the criterion for the test-taker's ability in order to determine the extent of the current test's validity.

2.2.2.4. Construct validity. Fulcher and Davidson (2007) claim that in order to internalize the construct validity it is needed to understand what a construct is. First, a construct should be measurable, and then it should be related to the separate constructs. Hughes (2003) states that to be able to determine the construct validity of a test, *think aloud* and *retrospection* techniques may work. In the former one, test-taker depicts their thoughts while answering the test. In the latter one, after the test finishes the test-taker tries to remember what they thought while responding.

2.2.2.5. Face validity. Brown (2004) states that if a test seems to be appropriate for what it is prepared to test, then it has face-validity. To give a simple example, an English language test prepared for young learners in no case should appear to be a Maths test for young learners. Underhill (1987) suggests for the test-makers in order to ensure face-validation, they get the ideas of the experts before administering it.

2.2.2.6. Response validity. Henning (1987) defines response validity as the validity that ensures the anticipated responses from the students. Some unexpected

problems may arise during a test administration and these problems negatively affect the response validity.

2.2.4. Authenticity

According to Bachman and Palmer (1996), a test can be labelled as authentic when it is in conformity to target language's real-word usage. This is something that is highly significant in young learners' language teaching: Since they are more familiar with the mother tongue acquisition in their real world, they need to have access to the language in an authentic way (Cameron, 2005). With regard to this point of view, teachers should increase the possibility of authentic language usage by providing students with authentic tasks in assessment (East, 2008).

2.2.5. Backwash Effect

Backwash or washback is the effect of tests on teaching and learning (Hughes, 2003; Heaton, 1990). It appears in two manners: *beneficial* (desirable) and *harmful* (undesirable) backwash effect. In order to achieve beneficial backwash, Hughes (2003, pp. 53-56) presents some advice for teachers:

> ➢ Test the abilities whose development you want to encourage.
> ➢ Sample widely and unpredictably.
> ➢ Use direct-testing.
> ➢ Make testing criterion-referenced.
> ➢ Base achievement tests on objectives.
> ➢ Ensure the test is known and understood by students and teachers.

Promodou (1995) points out why the beneficial backwash is hard to achieve lies in the teachers' attitudes toward the nature of testing; Teachers just concentrate on the tests' 'goodness' instead of the possible consequences of them on students and their learning.

2.3. Language Assessment Types

2.3.1. Traditional Language Assessment

Traditional assessment refers to standardized paper and pencil testing which focuses on the accurate production of structures in which all the students are expected to learn the same thing. Common item types in structure-based traditional assessment are summarized as follows (Simonson, Smaldino, Albright, and Zvacek, 2000):

> ➢ Multiple-choice items: practical items composed of a stem and a group of alternatives within those only one proves as the correct answer.

- ➢ True/False items: items which require students to decide whether the presented statement is true or not.
- ➢ Matching items: items which require student to detect and match two related written or visual element.
- ➢ Fill-in-the-blank items: items which include a missing part and require students to write the proper word or phrase in that part.
- ➢ Essays: effective assessment tools which require higher-order thinking skills and give the student the opportunity to write a paragraph or paragraphs freely around the required topic.

2.3.2. Communicative Language Assessment

Communicative Language Assessment first took place in language teaching field in 1980s. It took its roots from the theory of *communicative competence*. With a practical definition, communicative competence is "knowing when and how to say what to whom" (Larsen-Freeman and Anderson, 2011, p.115). In other words, communicative competence refers to the ability to use the language correctly and effectively in order to communicate in real life contexts. In 1970s the originator of communicative competence Dell Hymes was inspired by Chomsky's theory of linguistic competence which defines language ability as linguistic performance. In that sense, communicative competence entails more than linguistic knowledge of grammatical rules to communicate (Larsen-Freeman and Anderson, 2001).

A decade after the emergence of communicative competence, its influence on testing and assessment brought about Communicative Language Testing. Rather than learner's mere knowledge of language (grammatical competence), communicative testing embraces how the learner uses his/her receptive and productive skills in order to communicate in different social contexts (sociolinguistic competence), deal with communication breakdowns (strategic competence), and maintain the communication coherently (discourse competence) (Canale and Swain, 1980). According to Canale and Swain (1980, 1983) these skills constitute the communication itself and can never be fragmented.

Based on such a complex uniformity of linguistic elements, communicative language testing could not be implemented sufficiently by traditional paper and pencil tests (Clark, 1972; Oller, 1976). They lack the necessary components of communication such as authenticity, the performance and the context. Bachman and Palmer (2010) point out that

the content and the way of assessment are shaped by the content and the way of language instruction. With regard to these points of view, teachers are expected to organize not only their assessment but also their teaching taking a *whole* communicative teaching and learning into consideration. In this respect, four skills assessment should be in the foreground rather than structural assessment which is grammar-oriented.

2.3.3. Alternative Assessment

Shaaban (2005) points out that *alternative ways* of assessing students take into account variation in students' needs, interests, and learning styles; and they attempt to integrate assessment and learning activities. Also, alternative ways indicate successful performance, highlight positive traits, and provide formative rather than summative evaluation. Brown (2007) argues that the term *alternative assessment* brings about a misunderstanding in terms of its nature; instead he prefers to use *alternatives in assessment* indicating that tests are a part of the several possible alternatives, not outside of the responsible test formation. Common alternatives in assessment are portfolios, projects, self-assessment, peer assessment, journals, formal/informal observations, presentations, informal questioning, and teacher-student conferences inter alia.

Brown (2007, p.462) summarizes the most important characteristics of traditional and alternative assessments in Table 2.1.

Table 2.1. *Traditional vs. Alternative Assessment*

Traditional Tests	Alternatives in Assessment
One-shot standardized exams	Continuous long-term assessment
Timed, multiple-choice format	Untimed, free-response format
Decontextualized test items	Contextualized communicative tasks
Scores suffice for feedback	Formative, interactive feedback
Norm-referenced scores	Criterion-referenced scores
Focus on the "right" answer	Open-ended, creative answers
Summative	Formative
Oriented to product	Oriented to process
Noninteractive performance	Interactive performance
Fosters extrinsic motivation	Fosters intrinsic motivation

Brown (2007, pp. 475-479) defines common alternative types of assessment as follows:

➤ Portfolios: "purposeful collection of students' work that demonstrates to students and others their efforts, progress and achievements in given areas."
➤ Journals: written records such as "…language learning logs; grammar discussions; responses to reading; self-assessment; reflections on attitudes and feelings about oneself."
➤ Conferences: "a dialogue that is not to be graded" between teacher and student individually which aims to provide formative feedback to the student on any types of performance.

> ➢ Observations: "systematic, planned procedures for real-time, almost surreptitious recording of student verbal and nonverbal behavior" which aims to assess students at utmost degree without damaging the spontaneity of their performances.
> ➢ Self- and Peer Assessments: an autonomous assessment of students "to monitor his/her own as well as peers' performance and use the data gathered for adjustments and corrections".

As a matter of fact, many stakeholders agree on that traditional assessment is more practical than alternative assessment since alternative assessment requires more time, more subjective evaluation, more individualization, and more interaction in the process of providing feedback; However, the positive backwash effect of all these effort on alternative assessment makes it invaluable (Brown, 2004, 2007; Shaaban, 2005, Cameron, 2005, Mckay, 2006; Bachman and Palmer, 2010). Additionally, Cheng and Fox (2017, p.188) argue "teachers use assessment in their classrooms as something that is done *with* learners not *to* them" in order to stress the distinction between traditional and alternative types of assessment.

2.4. Testing and Assessing Young Learners

Before testing young learners' English language development, the first question is: Who are young learners? These learners of English as a second (L2) or foreign language (EFL) are defined in different age groups all over the world. A young learner is the child between the ages of five to seven in the very early years of school in Europe; while she/he is the learner between the ages of three to eleven spanning from pre-school to elementary school in USA (McKay, 2006; Nikolov, 2016; Shohamy, Or and May, 2017). In addition, Slattery and Willis (2001, p. 4) and Shin (2013, p.4) categorize young learners into two groups: "Very young learners: under the age of seven; young learners: between 7-12". In our country, the age range of young learners and consequently the student profile it refers to have changed several times since English became a compulsory subject in Turkish Education System in 1997. More detailed information about the history of English Teaching Program of Ministry of National Education (MoNE) is presented under the related title in this study. Currently the starting grade of English teaching is second grade (6-6.5 years); and the students are categorized as young learners until the age of 12.5 in Turkey (MoNE, 2018).

Besides the age range of the young learners, there are several other features differentiating these children from the adult learners of English. When it comes to the differences between adult and young learners, Cameron (2005) argues that significant differences come out of the linguistic, psychological and social development of the learners in addition to the more general features of children being more active, more

dependent on the teachers, more enthusiastic, more eager to do an activity even when they don't know anything about it (Cameron, 2001). Consequently, teachers of young learners need to take these characteristics into account while deciding the classroom activities, testing and assessment techniques, the body language they use and even the volume of their voices (Ionnaou-Georgiou and Pavlou, 2003; Nikolov, 2016) .

The characteristics of young learners, and the implications of these for the assessment of their language ability are a matter of discussion in the literature (Halliwell,1992; Vale and Feunteun,1995; Cameron, 2001; Rea-Dickins, 2000; Ionnaou-Georgiou and Pavlou, 2003). In line with the literature, Hasselgreen (2005) summarizes the common ground about the assessment principles in satisfying the following demands:

> • Tasks should be appealing to the age group, interesting and captivating, preferably with elements of game and fun.
> • Many types of assessment should be used, with the pupil's, the parents' and the teacher's perspectives involved.
> • Both the tasks and the forms of feedback should be designed so that the pupil's strengths (what he or she can do) are highlighted.
> • The pupil should, at least under some circumstances, be given support in carrying out the tasks.
> • The teacher should be given access to and support in understanding basic criteria and methods for assessing language ability.
> • The activities used in assessment should be good learning activities in themselves (pp. 338-339).

Another question for English teachers is: Why do we have to assess young learners of English? This question provides teachers the purpose of their assessment practices and it takes place in most of the works in language teaching and assessment field (Cameron, 2001; Hughes, 2003; Ioannou and Pavlou, 2003; McKay, 2006; Bachman and Palmer, 2010; Paker, 2013; Cheng and Fox, 2017). Upon the function of assessment in language education, Paker (2013) stresses the significance and necessity of assessment by pointing out that language skills cannot be undervalued by the students only when they are assessed. From this perspective, testing and assessment is an essential part not only of language teaching but also all the educational fields, for all the age levels of students, as well.

Since young learners are different from adults in nature, testing and assessing young learners may sound frightening. Nevertheless, it is an indispensable part of teaching and learning. In that case, one possible answer to this question may be that teachers want to be sure of the effectiveness of the teaching program and that children are really benefiting from the opportunity of learning a foreign language at an early age (Hughes, 2003). According to Hughes (2003), there is not an ideal test. A test may perfectly fit to an institution but turn in something useless for another one. In that manner, the purpose of the tests has great importance.

2.4.1. Assessing Four Skills of Young Learners

In the previous section, the questions of 'who are young learners' and 'why do we have to assess young learners' were focused on. In this section the question 'How do we assess young learners' four skills?' is the subject matter. This section provides some explanations on the strength of literature.

Children see the world from a colorful and enjoyable perspective. They can find joy in anything that never occurs to adults. They do not bother themselves with responsibilities or necessities. Hence they are not aware of the advantages or necessity of language learning until their parents or the school administration guide them to take language courses (Iannou-Georgio and Pavlou, 2003). That being the case, it is a must to make the language learning attractive to these young learners when they in any way happen to be a part of language education (Cameron, 2005; McKay, 2006; Nikolov, 2016; Cheng and Fox, 2017).

Ryan and Deci (2000) stress the *intrinsic motivation* for teaching young learners which occurs when the learners readily and eagerly engage in the task; conversely *extrinsic motivation* is outcome-oriented and lacks the enthusiasm of the learner. To be able to boost the intrinsic motivation of children, tasks and activities in which they have a voice when deciding to do should be proposed; by this way they gain self-confidence and desire to be involved in these tasks (Nikolov, 2016; Cheng and Fox, 2017).

As well as the way of assessment the tools which are involved in the assessment tasks are important. Shohamy et al. (2017) point out that "Choice of item and task types will need to correspond to the cognitive processing capabilities and degree of task familiarity of young learners" (p.9). Likewise, Cameron (2005) points out that children decide to go on or stop learning languages depending on the assessment and its backwash effect. Thus, it is worth keeping in mind while designing four skills assessment for young learners. All these specification are valid for both oral and literacy skills assessment. Based on the common grounds in literature, Cheng and Fox (2017) present a composition of tasks and assessment tools for four skills assessment of young learners. They are provided under the titles of skills in the following sections.

2.4.1.1. Assessing oral skills (listening & speaking). Listening skill involves some purposeful actions such as listening for the gist or detailed information. Young learners need to transfer their mother-tongue procedures to foreign language learning while listening to the target language such as inferring the meaning or guessing the content of the

listening input (Ioannou and Pavlou, 2003). In the same way, while engaging them in speaking, the things they are most familiar with such as their personal information or family can be included in the tasks. Compatible with the literature, the following are some assessment practices and tools suitable for both of the oral skills (Cheng and Fox, 2017, pp. 80-81):

1. Oral reading/dictation
2. Oral interviews/dialogues
3. Oral discussion with each student
4. Oral presentations
5. Public speaking
6. Teacher-made tests asking students to
a. give oral directions
b. follow directions given orally
c. provide an oral description of an event or object
d. prepare summaries of what is heard
e. answer multiple-choice test items following a listening passage
f. take notes
g. retell a story after listening to a recorded passage
7. Student portfolio
8. Peer-assessment / Self-assessment
9. Standardized speaking test
10. Standardized listening test

Over and above, all the suggested assessment procedures and tools require being applied considering young learners' characteristics and their progress in the language (Cameron, 2005; Ioannou and Pavlou, 2003). The reason is that "testing is more than a technical activity; it is also an ethical enterprise" (Fulcher and Davidson, 2007, p. xix).

2.4.1.2. Assessing literacy skills (reading & writing). Like oral skills, reading requires making use of some formative sub-skills such as reading in detail or reading for the gist. Teachers should be ready to practice and assess these skills regularly in order to improve young learners' reading skills. Burns and Siegel (2018) argue that reading is a valid tool in order to get input of the target language as well as listening, thus it requires teachers to provide that input to young learners; then assess them and provide feedback about their progress by applying the following assessment practices (Cheng and Fox, 2017, pp. 78-79):

1. Read aloud/dictation
2. Oral interviews/questioning
3. Teacher-made tests containing:
a. cloze items (e.g., words or phrases are systematically removed from a passage and students are required to fill in or identify what's missing)
b. sentence-completion items
c. true/false items
d. matching items
e. multiple-choice items
f. interpretative items (e.g. reading passage; or a map or a set of directions)

Writing requires the combination of other linguistic components such as grammar, vocabulary, syntax inter alia; thus teachers need to integrate writing with other skills and linguistic components by providing authentic context for young learners (Ioannou and Pavlou, 2003). Writing skill can be assessed by utilizing the assessment practices and tools as following (Cheng and Fox, 2017, pp.79-80):

1. Teacher-made tests containing:
a. true/false items
b. matching items
c. multiple-choice items to identify grammatical error(s) a sentence
d. editing a piece of writing such as a sentence or a paragraph
e. short essay
f. long essay
2. Student journal
3. Peer-assessment
4. Self-assessment
5. Student portfolio
6. Standardized writing tests

2.5. Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)

2.5.1. What is CEFR?

CEFR stands for an international framework for language learning, teaching, and assessment across Europe created by the Council of Europe (CoE) in 2001. The term "framework" refers to a common guideline aiming to provide a means for stakeholders in language teaching including teachers, in terms of all related components of language teaching, learning, and assessment, flexible and suitable for a variety of societies and languages (CoE, 2001; Trim, 2011; Arıkan, 2015). Trim (2011, p.2) declares about CEFR that "it was envisaged primarily as a planning tool whose aim was to promote 'transparency and coherence' in language education". In the light of the research on CEFR, it is obvious that CEFR is a pathfinder which suggests possible ways to adopt syllabuses, curriculum guidelines, strategies, examinations and materials not for a unique language but for all the European languages targeted to use for communication for the sake of *plurilingualism* and *pluriculturalism* (CoE, 2017). The addressees are anybody who is responsible from the insurance of meeting learners' needs to develop communicative competence. The descriptors of CEFR are presented as suggestions rather than compulsory actions (CoE, 2018). For that reason, English teachers need not only to internalize but also to adopt the objectives of CEFR in accordance with their own context and implement them in the classroom in an effort to make the students the *users of the language*, instead of just learners of it.

CEFR involves such huge amount of information related to language teaching, learning, and assessment that in literature a certain amount of research was conducted to serve as guidelines to understand and implement it in all dimensions (Fulcher and Davidson, 2007; Trim, 2011; Benigno, 2016). In the current study, the issues in CEFR, which are related to the aim of the study, are presented briefly.

2.5.2. Key Aspects of CEFR.

2.5.2.1. The aims of CEFR. CEFR was created to present a common base for language teaching and learning across Europe. In addition to that general purpose of its construction, it has some specific aims which are listed as follows:

- promote and facilitate co-operation among educational institutions in different countries;
- provide a sound basis for the mutual recognition of language qualifications;
- assist learners, teachers, course designers, examining bodies and educational administrators to situate and co-ordinate their efforts (CoE, 2001, pp. 6-7).

Over and above, according to CEFR (2001) the language learner is the *social agent* who uses the language and takes part in his/her own learning process actively. By this way the learner-autonomy is enhanced.

2.5.2.2. The Action-oriented approach and communicative competence. It is clearly stated in the CEFR that it doesn't favor any kind of approach or prevent the others. Instead, it takes a comprehensive position in which all the possible approaches are suggested with regard to the learners' needs. The action-oriented approach requires a transition from the pre-created curriculums to the curriculums which take the learners' needs in the center. It cares for what the learner is able to do instead of the inabilities of the learner; and facilitates the abilities by providing real-life tasks (CoE, 2018).

As for the communicative competence it is the combination of the abilities of using the language linguistically correct and coherent (linguistic competence), socially appropriate (socio-linguistic competence), and justifying the continuity of the communication in defiance of communication breakdowns (pragmatic competence) (Hymes, 1966; CoE, 2001, 2018). CEFR states that the learners when exposed to the language in their daily lives need to use it and develop communicative language skills which contribute to their communicative competence. Furthermore, it is remarked in CEFR that communication is supported by communicative tasks of which the purpose is determined with regard to the needs in a given situation (CoE, 2001, 2018).

2.5.2.3. CEFR common reference levels. CEFR presents three broad levels (A, B, and C) and their sub-levels (A1, A2, B1, B2, C1, C2). A level proficiency refers to the 'basic user'; B level proficiency refers to the 'independent user'; and C level proficiency refers to the 'proficient user' (See Table 2.2. *Common Reference Levels, CoE, 2001*). Additionally, CEFR recommends more sub-levels for the proficiencies such as A2+, B1+ when necessary.

In Turkey MoNE (2018) adopts the A level proficiency for young EFL learners as basic users of EFL. From the 2nd grade up to the 6th grade equals to A1 level while 7th and 8th grades equals to A2 level. CoE (2001, p.24) presents a global scale for the general proficiency levels (See Table 2.2.).

Table 2.2. *Common Reference Levels: Global Scale (CoE, 2001, p.24)*

Proficient User	C2 (Mastery)	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1 (Effective Operational)	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices.
Independent User	B2 (Vantage)	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1 (Threshold)	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2 (Waystage)	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1 (Breakthrough)	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

In addition to this general proficiency levels, CEFR provides proficiency levels for each skill. According to these levels, reading and listening constitute the skills of *understanding;* speaking is divided into two categories as *spoken interaction and spoken production*; and writing skill is the same. Another elaboration for spoken language in CEFR is presented as "qualitative aspects of spoken language use" with proficiency levels of "range, accuracy, fluency, interaction, and coherence" (CoE, 2001, pp. 28-29).

2.5.2.4. Concepts related to language assessment and European language portfolio. Some concepts elaborated in the first chapter of this study are reverted in this section with the interpretation of CEFR. CEFR (CoE, 2018) highlights the relationship between test, assessment, and evaluation. *Tests* are the tools in the form of written materials requiring students to show their knowledge and used to assess students' linguistic performance. *Assessment* is a broader concept than test since there are assessment tools which are not tests. On the other hand, *evaluation* is a way to determine the effectiveness or quality of other proponents of education such as the program, curriculum, and materials and inter alia rather than the student proficiency. In that sense, assessment is a part of evaluation which is the broadest one within these terms.

CEFR stresses the significance of another three concepts in language assessment: *validity, reliability, and feasibility. Validity* is the term which demonstrates whether the test is able to assess what it is intended to. *Reliability* is a term which indicates extent of the rank order which proves the same results when the same assessment administered twice. Finally, *feasibility* refers to the terms *practicality*, mentioned in the first chapter. Feasibility is related to the possibility to achieve the performance in a reasonable period, with a reasonable effort (CoE, 2001, 2018).

According to CEFR (CoE, 2001, pp. 183-191) assessment can be categorized in plenty of ways:

> *Achievement assessment:* the assessment of the achievement of specific objectives – assessment of what has been taught. It therefore relates to the week's/term's work, the course book, the syllabus.
>
> *Proficiency assessment:* assessment of what someone can do/knows in relation to the application of the subject in the real world. It represents an external perspective.
>
> *Norm-referencing (NR):* the placement of learners in rank order, their assessment and ranking in relation to their peers.
>
> *Criterion-referencing (CR):* Reaction against norm-referencing in which the learner is assessed purely in terms of his/her ability in the subject, irrespective of the ability of his/her peers.
>
> *Mastery Criterion Referencing:* One in which a single 'minimum competence standard' or 'cut-off point' is set to divide learners into 'masters' and 'non-masters', with no degrees of quality in the achievement of the objective being recognized.

Continuum Criterion Referencing: An approach in which an individual ability is referenced to a defined continuum of all relevant degrees of ability in the area in question.

Continuous assessment: Assessment by the teacher and possibly by the learner of class performances, pieces of work and projects throughout the course. The final grade thus reflects the whole course/year/semester.

Fixed assessment points: When grades are awarded and decisions made on the basis of an examination or other assessment which takes place on a particular day, usually the end of the course or before the beginning of a course. What has happened beforehand is irrelevant; it is what the person can do now that is decisive.

Formative assessment: An ongoing process of gathering information on the extent of learning, on strengths and weaknesses, which the teacher can feed back into their course planning and the actual feedback they give learners.

Summative assessment: Sums up attainment at the end of the course with a grade. It is not necessarily proficiency assessment. Indeed a lot of summative assessment is norm-referenced, fixed-point, achievement assessment.

Direct assessment: Assessing what the candidate is actually doing. For example, a small group is discussing something, the assessor observes, compares with a criteria grid, matches the performances to the most appropriate categories on the grid, and gives an assessment.

Indirect assessment: Uses a test, usually on paper, which often assesses enabling skills..

Performance assessment: Requires the learner to provide a sample of language in speech or writing in a direct test.

*Knowledge assessment: R*equires the learner to answer questions which can be of a range of different item types in order to provide evidence of the extent of their linguistic knowledge and control.

Subjective assessment: A judgement by an assessor. What is normally meant by this is the judgement of the quality of a performance.

Objective assessment: Assessment in which subjectivity is removed. What is normally meant by this is an indirect test in which the items have only one right answer, e.g. multiple choice..

Checklist rating: Judging a person in relation to a list of points deemed to be relevant for a particular level or module.

Impression: Fully subjective judgement made on the basis of experience of the learner's performance in class, without reference to specific criteria in relation to a specific assessment.

Assessment by others: Judgements by the teacher or examiner.

Self-assessment: Judgements about your own proficiency.

Another important element of global assessment suggested by CEFR is the *European Language Portfolio.* According to CEFR, "The European Language Portfolio (ELP) provides a format which makes it possible for learners to document their progress towards plurilingual competence by recording learning experiences of all kind over a wide range of languages" (CoE, 2001, p.5). The ELP help learners be aware of their regular improvements in learning languages by self-assessing themselves (CoE, 2018). It is recommended not just for the adult language learners but also for the young and very young learners of all second and foreign languages.

2.6. Ministry of National Education's English Language Teaching Program for Primary and Secondary Schools (2nd, 3rd, 4th, 5th, 6th, 7th and 8th Grades)

2.6.1. Historical Changes in the Curriculum

In the following paragraphs, a brief history of educational reforms in Turkish education system and ELT curriculums made by MoNE and the changes in testing and assessment implementations are summarized. Additionally, testing and assessment practices, tools and techniques written in the ELT Program suggested for young learners by MoNE are explained.

While the education policies are determined, countries' scientific, economic, political, social and cultural needs have a great influence on the decisions about foreign language education. Early language learning was not a compulsory component of primary school curriculum in Turkey until the educational reform in 1997. In 1997 reform, the Turkish Ministry of National Education (MoNE) and Higher Education Council (HEC) made dramatic changes in both educational system and English Language Teaching (ELT) policy (Kırkgöz, 2007). With this reform, the duration of compulsory education was increased to eight years from five years; in addition, ELT in 4th and 5th grades became compulsory (Doğançay-Aktuna and Kiziltepe, 2005; Gürsoy, Çelik Korkmaz, and Damar, 2013). Communicative Language Teaching (CLT), learner-centeredness, teacher's role as a facilitator, learner autonomy, and contextual teaching was firstly introduced by the 1997 curriculum (Damar, 2004; Gürsoy et al., 2013). By means of CLT, the four communication skills listening, speaking, reading and writing were firstly involved in the ELT Program of MoNE (Kirkgöz, 2007; Arslan 2012).

Until the revision of the 1997 curriculum in 2005, testing and assessment in EFL classrooms had been based on traditional structure-based paper-and-pencil tests. In that year Turkey revised the EFL teaching and assessment based on CEFR (Kirkgöz, 2007; Arslan 2012). After that revision, performance-based assessment, based on European Language Portfolio (ELP), was proposed through implementing portfolios in parallel with the principles of CLT (Kirkgöz, 2007).

The scientific and technological developments in recent years have had great influence in foreign language education as well as the other areas. Thus, Turkey went through a new educational reform named 4+4+4 educational reform which is a transition from 8+4 educational model in 2012. The new curriculum offered considerable changes in

language education as well as the three-tier system. The starting age for schooling has been lowered to 5-5.5 years of age from seven: and starting age for the foreign language learning has been lowered to 6-6.5 years of age (2nd grade) from nine (MoNE, 2013, 2015, 2018) (See Table 2.3).

In 2017, MoNE took an action to revise the ELT Program with the aim of placing the *values education* into the new curriculum. In this revision, English teachers, parents, and academicians all over the country were consulted through written documents of their views to make recommendations and critiques for the available curriculum in order to revise it with the participants' confirmation (MoNE, 2017). After that, MoNE published the sample ELT Program before the beginning of 2017-2018 academic year. Finally, the new revised version of the ELT Program for Primary and Lower-secondary Schools were published in 2018. This new version of the ELT Program consisted of not the dramatic changes but the updated forms of the curricular components. The major dimensions of the revision on the new curriculum were presented as follows (MoNE, 2018, p. 3):

> 1. Revision of the theoretical framework;
> a. Reviewing the curriculum with regards to values education
> b. Including the basic skills as themes
> c. Expanding certain subsections, such as testing and evaluation, and suggestions
> 2. Revision of each grade by;
> a. Revision of the target language skills and their linguistic realizations
> b. Evaluation and the update of the contexts, tasks and activities
> c. Analysis and general update of the curriculum in terms of functions and forms covered".

In the new version of the ELT Program, it is pointed out that this new version differs from the previous one by means of the emphasis on "the values education". Values reflect the characteristics of our society as an indispensable part of it (Yazıcı, 2014). For this reason it is aimed that our children, the guarantee of our country's future, grow with this conscious (MoNE, 2018). The featured values in the program are: "friendship, justice, honesty, self-control, patience, respect, love, responsibility, patriotism, and altruism" (MoNE, 2018, p. 6). MoNE also puts the emphasis on the integration of these values in the themes and topics rather than presenting them as separate subjects.

The ELT Program of MoNE is based on the principles of Common European Framework of Reference for Languages: Learning, Teaching, and Assessment (CEFR) which adopt the communicative competence as basis in language teaching (MoNE, 2013, 2018). In the following sections of the study a detailed expression of CEFR is presented. The ELT Program has been axes: instructional design, instructional materials, and assessment in language teaching (MoNE, 2013, 2018). Within these axes, focus of the current study is on the assessment in language teaching.

The new curriculum emphasizes listening and speaking over literacy skills (reading and writing) for the A1 level young learners since they are not familiar with the written and oral communication and need to be exposed to the language orally. For this reason, in the 2nd grade literacy skills (reading and writing) are not included; they are introduced at the 3rd grade at word basis. From the 5th grade and so forth the percentages of the literacy skills increase; however, they stand behind the oral skills (listening and speaking) even at the A2 level. Until 2017, the total length of language teaching per week was limited to two hours at A1 level in primary schools, and four hours at both A1 and A2 level in lower-secondary schools (see Figure 2.2.). In addition, schools administrations were free to add two hours of elective English courses at lower-secondary schools.

Levels [CEFR*] (Hours / Week)	Grades (Age)	Skill focus	Main activities/ strategies
1 [A1] (2)	2 (6-6.5)	Listening and Speaking	TPR/Arts and crafts/Drama
	3 (7-7.5)	Listening and Speaking Very Limited Reading and Writing°	
	4 (8-8.5)	Listening and Speaking Very Limited Reading and Writing°	
2 [A1] (4)	5 (9-9.5)	Listening and Speaking Limited Reading° Very Limited Writing°	Drama/Role-play
	6 (10-10.5)	Listening and Speaking Limited Reading° Very Limited Writing°	
3 [A2] (4)	7 (11-11.5)	**Primary:** Listening and Speaking **Secondary:** Reading and Writing	Theme-based ∞
	8 (12-12.5)	**Primary:** Listening and Speaking **Secondary:** Reading and Writing	

Figure 2.2. Model English Language Curriculum (MoNE, 2015, p.V)

Figure 2.2 demonstrates the total lengths of English courses per week in the 2015 version of the ELT Program of the MoNE. In 2017, MoNE modified the total lengths of English must courses in lower-secondary schools. At 5th and 6th grades it was reduced to three hours per week instead of four hours while the total hours of English must course did not change at 7th and 8th grades. (See Table 2.3.). Additionally in the new schedule, at lower-secondary schools there are two hours of elective English courses. Since then, this schedule has been implemented at state primary and lower-secondary schools.

Table 2.3. *Weekly Corse Schedule for 2nd-8th Grades (MoNE, 2018)*

	2nd G.	3rd G.	4th G.	5th G.	6th G.	7th G.	8th G.
English Must Course (hours/week)	2	2	2	3	3	4	4
English Elective Course (hours/week)	-	-	-	2	2	2	2

Note: It is retrieved from http://ttkb.meb.gov.tr/www/haftalik-ders-cizelgeleri/kategori/7

In 2018 MoNE regulated the number of achievement exams in primary schools. It was declared that at all grade levels two EFL exams would be administered regardless of the duration of English lessons. Before that regulation, three exams had been administered for more than three hours of English lessons in a week at each grade level.

2.6.2. Testing and Evaluation Approach of the ELT Curriculum based on CEFR

Following the recent changes in the ELT program, certain key concepts have taken considerable amount of emphasis. *Learner autonomy, problem-solving skills, communicative competence, authentic sources, and action-oriented approach* stand for the basis of the structure of ELT Program. Parallel with the CEFR, the revised version of ELT declares that language is a way of communication, not a subject to learn just in the school settings. From that point of view, it is emphasized that testing procedures and the nature of the curriculum should be consistent with each other. Otherwise, all the elements of educational and linguistic structure and the objectives are put on the line (MoNE, 2018). In line with this principle MoNE (2018) stresses the unity of teaching, learning and assessment in the sense that it determines the teachers' and students' strategies as well as the parents' attitudes towards the values in education. It is stated that in order to create *beneficial backwash effect* on the whole teaching and learning process, teaching and assessment is based on the same objectives (MoNE, 2018). After the 2012 reform in the Turkish educational system and the ELT Program, MoNE (2013, p. XV) suggested four types of assessment (See Figure 2.3.).
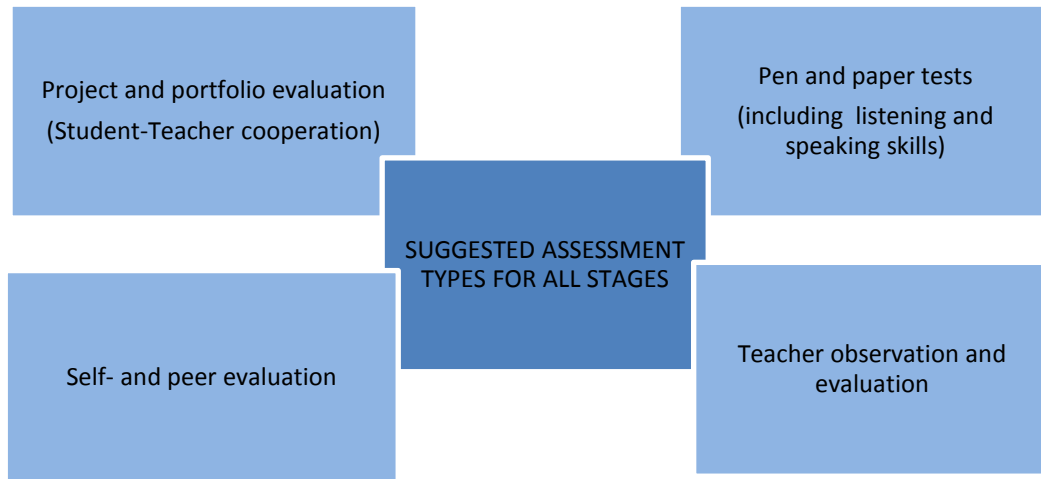
Figure 2.3. Suggested Assessment Types for All Stages (in 2012 reform)

The suggested assessment types were limited to four types of assessment and they were presented superficially in terms of implementation.

After the revision in 2018, MoNE made dramatic changes in the suggested assessment types and techniques. When compared to the previous types of assessment, this new version has a broader scope and proves more relevant to CEFR in terms of diversity taking the students' needs into account (see Table 2.4.). The revised curriculum (2018) supports a mixture of all assessment types instead of overuse of certain assessment techniques. In the new ELT Program, "learner autonomy" and "communicative competence" in language teaching has certain emphasis. In that sense, *self-assessment, alternative, and process-oriented assessment* are within the main suggested assessment tools. For the implementation of self-assessment technique, in the English course books, there are check-lists for students involving questions about their own language learning progress such as "What did you learn?", "How much do you think you learned?" and "What do you think you can do in real life, based on what you learned in class?" (MoNE, 2018, p. 6).

Table 2.4. *Suggested Testing Techniques for the Assessment of Four Skills (adapted from the ELT Program of MoNE (2018, pp. 7-8))*

Language Skills	Testing Techniques
Speaking	Collaborative or singular drama performances (Simulations, Role-plays, Side-coaching), Debates, Group or pair discussions, Describing a picture/video/story, etc., Discussing a picture/video/story, etc., Giving short responses in specific situations, Information gap, Opinion gap, Reporting an event/anecdote, etc., Short presentations, Talking about a visual/table/chart, etc.
Listening	Different variations of matching (…the sentences with paragraphs … pictures with the sentences, etc.), Discriminating between phonemes, Identifying interlocutors' intentions and implicatures, Listen and perform/complete an action (E.g.: Listen and draw/paint, listen and match, listen and put the correct order, listen and spot the mistake, etc.), Listen and tick (the words, the themes, the situations or events, the people, etc.), Omitting the irrelevant information, Putting into order/reordering, Recognizing phonemic variations, Selective listening for morphological structure and affixation, True/False/ No information, Understanding overall meaning and supporting details, Recognizing specific information, Questions and answers
Reading	Different variations of matching (…the sentences with paragraphs, … pictures with the sentences, etc.), Finding specific information, Finding a title to a text, Identifying the gist and supporting details, Intensive reading, Read and perform / complete and action (E.g.: Read and guess the meaning of lexemes, Read and draw/paint, Read and solve the riddle), Solving a puzzle, Spotting text mechanics (reference, substitution, various types ellipses), True/False/No information, Transferring the text to a table/chart (Information transfer), Understanding the author's intention, Questions and answers
Writing	Describing a picture/visual/video, etc., Filling in a form (hotel check in form, job application form, etc.), Note taking/ making, Preparing an outline, Preparing a list (shopping list a to-do list, etc.), Reporting a table or a chart, Rephrasing, Rewriting, Writing short notes, entries and responses, Writing a paragraph/e-mail/journal entry/etc., Writing a topic sentence/ thesis statement
Samples for integrated Skilss	Summarizing a text (listening/reading and writing), Taking notes (listening and writing), Reporting an event (listening/reading and speaking), Paraphrasing (listening/reading and writing), Preparing a mind-map (reading/ listening and writing), Cloze/C-test (reading and writing), Dictation (listening and writing), Reading a text and present it (reading and speaking), Writing a text and present it (writing and speaking), Outlining a reading text (reading and writing)
Alternative Assessment	Portfolio Assessment, Project Assessment, Performance Assessment, Creative Drama Tasks, Class Newspaper/Social Media Projects, Journal Performance, etc.

On the other hand, formal assessment tools such as written and oral exams, quizzes, homework and projects are within the suggested assessment tools in the new ELT Program (MoNE, 2018). It is stated in the new ELT Program (MoNE, 2018) that 2nd and 3rd graders are not assessed via summative testing tools; instead the young learners in these grades are suggested to be assessed with the help of formative procedures. However, the young learners in the 4th grades and the lower-secondary grades (5th, 6th, 7th, and 8th grades) are

suggested to be assessed via both summative and formative assessment tools and techniques in both product and process-oriented procedures.

The purposes of all the assessment tools suggested by MoNE (2018, p.7) are listed as follows:

> • cover four language skills and implicit assessment of language components;
> • vary in terms of learning styles and cognitive characteristics of the students;
> • be in consistent with the learning and teaching methodology depicted in the curriculum;
> • be in line with the students' developmental characteristics;
> • create positive and beneficial washback effect;
> • include self-assessment, reflection and feedback and
> • help students identify their strengths and weaknesses and target areas that need work.

In the light of the purposes of the assessment procedures above, it is obvious that MoNE (2018) suggests EFL teachers of young learners to utilize all the possible assessment techniques and tools having regard to young learners' developmental features and the objectives of the courses based on communicative competence, including the four language skills, as well. Considering this point of view, this study aims to find out to what extent the EFL teachers of young learners working in state lower-secondary schools follow the suggested procedures while assessing young learners.

2.7. Research Studies on Consistency between Curriculum and Implementation in Language Assessment and Assessment of Young Learners' EFL

Pandian (2002) wrote an article on a shift from grammar-based to communication-based teaching on language education in Malaysia. In the article, Pandian reports a study conducted by the curriculum development centre (Berita Kurikulum, 1999) in Malaysian context. Several high-achieving rural primary schools participated in the study. The findings of the study revealed that in classroom-teaching "chalk-and-talk drill method" (Pandian, 2002, p.42) was applied instead of communicative competence. Furthermore, the findings indicated that in the sixth grade, grammar, reading, writing and vocabulary were in the foreground rather than communicative skills such as listening and speaking in Malaysian primary schools. The underlying reason was that listening and speaking did not take place in the primary educational institutions' assessment programs. Pandian also points out that the ministry has taken some innovative actions to avoid being stuck to structural language education in Malaysia.

Yang (2008) investigated the assessment types used by the Taiwanese primary school EFL teachers. The aim of the study was to find out how frequently traditional and

alternative assessment procedures were applied in young learners' classrooms in Taiwan. 425 EFL teachers from several cities participated in the study. Data were gathered by a questionnaire. Results revealed that multiple choice and matching items were among the most frequently preferred traditional assessment tools by the teachers while fill-in-the blank and true/false items followed multiple-choice and matching items. On the other hand, teachers used short-answer items, restricted-response essays, and extended-response essays least frequently. Among the alternative assessment tools, teachers mostly used informal questioning, musical presentations, and informal observations and role-plays while the least frequently used alternative assessment tools were journals and creative writing. The mean scores of alternative assessment tools were higher than the traditional assessment tools. The researcher interpreted that EFL teachers made use of both traditional and alternative assessment tools in their classes with different frequencies.

Brumen, Cagran, Coombe, Edmonds, Heckstall-Smith and Fleming (2009) conducted a study named "Comparative Assessment of Young Learners' Foreign Language Competence in Three Eastern European Countries" with the aim of learning EFL teachers' ideas and practices in the assessment of young learners. The participants were 108 EFL teachers from Slovenia (50 teachers), Croatia (28 teachers), and Czech Republic (30 teachers). Data were collected by an adapted questionnaire. Results showed that Croatian teachers were prone to assess listening and speaking more frequently than the other teachers from the two other countries did. On the other hand, Czech teachers mostly went in for assessing the literacy skills (reading and writing) rather than oral skills (listening and speaking). Finally, Slovenian teachers tended to use more grammar and vocabulary-oriented tests when compared to the other teachers. Other findings of the study were related to the item types used in the tests. It was found out that Croatian teachers made use of repeat-and-drill practices of vocabulary in most cases while Slovenian teachers preferred fill-in-the blank type of activities. On the other hand, Czech teachers used mostly activities including true/false type of items.

Yildirim and Orsdemir (2013) conducted a study named "Performance Tasks as Alternative Assessment for Young EFL Learners: Does Practice Match the Curriculum Proposal?" In this study the emphasis was on the performance tasks as an alternative assessment tool. The researchers aimed to find out the reality in classrooms of young learners in terms of availability of performance tasks in line with the Policy's proposals. 43 EFL teachers who had experience in the ELT field more than ten years participated in the study. Data collection instruments were a questionnaire, interviews with teachers and

document analysis of the performance assessment tools. Results revealed that in their responses to the questionnaires and interviews, teachers declared that they utilized performance tasks effectively compatible with the curriculum. However, the document analysis displayed a different scene in which listening, reading and speaking skills were totally ignored while writing and grammar were slightly fostered. Additionally the allocated performance tasks were not proper for improving higher-order thinking skills; rather simple psycho-motor skills were improved to some extent in preparation of the performance materials. Thus, the researchers concluded that rather than a match, a mismatch showed up between EFL teachers' assessment practices of performance assessment in young learners' classrooms and the policy's proposals.

Han and Kaya (2014) conducted a survey with the aim of finding out Turkish EFL teachers' perceptions and in-class practices of assessment in terms of four skills assessment in the context of constructivist instruction. In addition, teachers' preferences of EFL assessment in terms of its purpose were intended to be found out. The participants were 95 EFL teachers working in primary and secondary state schools in Turkey. The data was conducted via questionnaires. Han and Kaya (2014, p.82) adopted "Classroom Assessment Preferences Survey Questionnaire for Language Teachers (CAPSQ-LT) (Gonzales and Aliponga, 2012)". The assessment preferences presented in the survey were "assessment as learning, assessment of learning, assessment for learning, assessment for instruction, and assessment to inform" (Han and Kaya, 2014, p.83) The results revealed that most teachers perceived assessment as learning. Namely, they perceived assessment as a tool to foster the students' learning. On the other hand, reading and writing skills were mostly assessed by EFL teachers while listening and speaking were assessed less frequently. In addition, teachers thought that assessing speaking was by far the most demanding work while reading assessment was the least demanding among four skills assessment. Furthermore, teachers assessed speaking skill mostly through role-plays; reading skill through true/false and multiple choice items, listening skill through dialogues, and writing skill through writing about past events. There were not significant differences between male and female teachers' practices of EFL assessment. Furthermore, teachers did not change their preferences of assessment whether they received pre-service and in-service training of assessment.

Basok (2017) investigated the consistency between policy and implementations of the curriculum by the EFL teachers in Turkey. Three Turkish EFL teachers were interviewed in thirty-minute periods about the policy's EFL assessment suggestions and

the actual situation in Turkish primary, secondary and high school settings. Results of the study indicated that teachers felt under pressure by the language examinations administered by the government which the students were to take to be able to enter a higher level of educational institution. For that reason teachers could not implement what the policy suggested. Instead, they preferred to prepare the students for the examinations by using grammar-based teaching and assessment practices. Additionally, primary school EFL teachers declared that they felt more flexible in implementing the policy's suggested communicative teaching and assessment procedures since their students were not required to take high-stake language examinations.

CHAPTER III: METHODOLOGY

This chapter presents the methodology applied in the study to investigate EFL teachers' testing and assessment practices in lower-secondary schools and their consistency to the ELT Program of MoNE in Turkish context. It covers the research methods applied in the analysis of the data obtained, the research design which elaborates on the features of the determined research methods and how they work, the setting and the participants of the study, the data collection tools as instrumentation, the data collection procedure, and the analysis of the data.

3.1. Research Methods

In Social Sciences, there are two main research types: *quantitative and qualitative research methods.* Quantitative research method yields results from the numerical, countable forms of data turning into statistical analysis; qualitative data yields results from verbal data obtained from observations, interviews or documentaries and requires a subjective analysis and interpretation (Creswell, 2014; Edmonds and Kennedy, 2017; Cohen, Manion and Morrison, 2018). However, when there are gaps or deficiencies in either of the methods, a mixture of them which brings both quantitative and qualitative elements together in one study is required. This new mixture is called *mixed-methods* (Creswell, 2014; Cohen, et al, 2018). Creswell (2018, p.17) illustrates the interrelationship within these research methods in Table 3.1.

Table 3.1. *Quantitative, Mixed, and Qualitative Methods*

Quantitative Methods	Mixed-Methods	Qualitative Methods
Pre-determined	Both pre-determined and	Emerging methods
Instrument-based questions	emerging methods	Open-ended questions
Performance, attitude, observational and census data	Both open and closed-ended questions	interview, observation, document, and audiovisual data
Statistical analysis	Multiple forms of data	Text and image analysis
Statistical interpretation	Statistical and text analysis Across-databases interpretation	Themes, patterns interpretation

Mixed-methods research is based on pragmatism. On account of pragmatism, mixed-methods research involves non-rigid data collection procedure which encourages the researcher to be free to combine quantitative and qualitative methods in a unique study (Creswell, 2003, 2014). Researchers facilitate their studies with the help of mixed-methods research when they have difficulty in finding answers to their research questions by

numbers in quantitative method and by verbal data in qualitative method (Creswell, 2014; Cohen et al, 2018).

This study seeks to examine EFL teachers' testing and assessment practices and their consistency to ELT Program suggested by MoNE. In the light of literature, this study was designed as a mixed-methods research. Single application of questionnaire as a means of quantitative data would result in insufficient information about how the EFL teachers assess lower-secondary school students' English since assessment practices involve real life actions of teachers and students. On the ground that assessment documents used by the EFL teachers in real classrooms would enhance the results of quantitative data as a means of qualitative data, the researcher designed this study as mixed-methods research.

3.2. Research Design

Research design reflects the characteristics of the research methods which enable the researcher to decide how to collect, analyze and interpret the data of the study. Each research method has different types of research designs (Creswell 2003, 2014; Cohen et al., 2018).

Deciding the type of the method as mixed-methods research in this study, next step is to decide the priority of the methods which come together in mixed-methods. In order words, whether the quantitative method or the qualitative method would take the priority or both would work equally (Edmonds and Kennedy, 2017). In literature the priority of the methods is symbolized by capitals in a short form as QUAN for quantitative QUAL for qualitative method (Creswell, 2003, 2014; Edmonds and Kennedy, 2017; Cohen et al., 2018). Thus in this study it is explained in the same form. Another aspect to decide is the scheduling of the data which determines the sequence of collection. Lastly, it is necessary to decide whether the data would be analyzed separately or together. The decisions made on the basis of aforementioned aspects of the study generate the type of mixed-methods research (Creswell, 2003, 2014; Edmonds and Kennedy, 2017).

In this study, *parallel databases design* under the *Convergent Parallel Approach* of mixed-methods research was applied. "This approach involves the collection of different but complementary data on the same phenomena. Thus, it is used for the converging and subsequent interpretation of quantitative and qualitative data" (Edmonds and Kennedy, 2017, p.181). Parallel databases design involves the simultaneous but separate collection of the quantitative and qualitative data. "This design allows researchers to validate data by converging the QUAN results with the QUAL findings" (Edmonds and Kennedy, 2017, p.182). Edmonds and Kennedy (2017)

explain the priorities of the methods by capital letters: when both of them in the capitals it means both of them have equal priority in the study.

In this study, the quantitative method provides information about the teachers' preferences of item types in terms of traditional and alternative assessment types with the help of a questionnaire. Qualitative method provides the information about how frequently the EFL teachers assess four skills in two separate measures: 1. Open-ended questions in the questionnaire; 2. Assessment documents of teachers used in the real classrooms. The quantitative and qualitative data were collected at the same time but with the help of different collection tools in conformity with parallel databases design.

3.3. Setting and Participants

3.3.1. Setting

This study aims to find out the testing and assessment practices of EFL teachers working with young learners at 5th, 6th, 7th and 8th grades. With that purpose the target population of the study was EFL teachers working in state lower-secondary schools in two central districts (Merkezefendi and Pamukkale) in Denizli. Researcher intended to reach as many EFL teachers as possible to be able to reflect the whole picture of teachers' assessment practices. With that purpose, information about names and number of the state lower-secondary schools as well as the number of the target EFL teachers was obtained from the Provincial Directorate of National Education. Giving the priority to the schools which hosted more EFL teachers, nearly all of the state lower-secondary schools in Merkezefendi and Pamukkale were visited in order to interact with the EFL teachers at the end of the spring term of 2017-2018 academic year.

3.3.2. Participants

In the target districts of Denizli there were 286 EFL teachers working in 70 state lower-secondary schools in the spring term of 2017-2018 academic year. On a voluntary basis, 152 EFL teachers out of 286 accepted to participate in the study. All of the 152 teachers participated in the quantitative data collection procedure of the study. However, in the qualitative data collection procedure 41 out of 152 teachers voluntarily shared their documents they used in testing and assessing students.

In most cases the demographical features of the participants affect their tendencies or preferences in certain situations. This idea in mind, one of the sub-questions of the research seeks to find out whether the teachers' gender, duration of experience and the

highest degree they hold academically affect their preferences of traditional and alternative assessment types or not. Figure 3.1. shows the distribution of EFL teachers' gender.
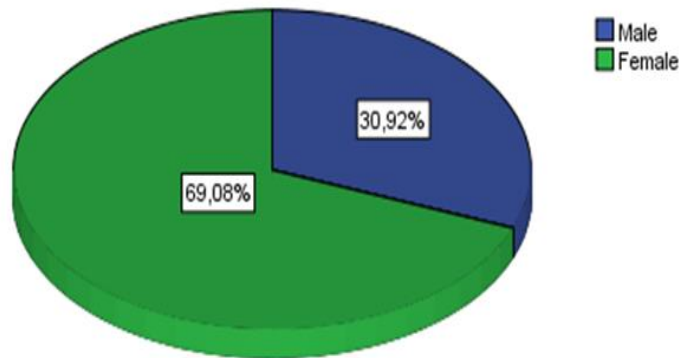
Figure 3.2. Distribution of Gender of the EFL Teachers

As the percentages are demonstrated in Figure 3.1., 69.08 % of the participants were female EFL teachers while the 30.92 % of the participants were male EFL teachers. Table 3.2 demonstrates faculties and departments which teachers hold their Bachelor of Arts.

Table 3.2. *Techers' Bachelor of Art (BA).*

Faculty /Department	Frequency	Percentage %
Faculty of Education / Department of ELT	115	75.7
Open Education Faculty / Department of ELT	2	1.3
Faculty of Science and Literature /Department of English Language and Literature	17	11.2
Faculty of Science and Literature / Department of American Culture	2	1.3
Faculty of Science and Literature / Department of English Linguistics	1	0.7
Faculty of Science and Literature / other	3	2.0
Other teaching branches (Maths teaching, Turkish teaching, Primary school teaching, Science teaching etc.)	11	7.2
Another department which is not directly related to English or teaching (Physics, Chemistry, Engineering etc.) Chemistry	1	0.7
Total	152	100.0

Table 3.7 gives detailed information about the faculties and departments where the teachers graduated. According to the table, teachers graduated from several faculties and departments. However, the vast majority (75.7 %) of teachers graduated from Department of ELT. 17 teachers (11.2 %) graduated from English Language and Literature. 11 teachers graduated from other teaching branches such as Maths teaching, Turkish teaching and Primary School teaching etc. For teachers' duration of experience see Table 3.3.

Table 3.3. *Experience of the Teachers*

Teaching experience	Frequency	Percentage
1-5 years	16	10.5
6-10 years	48	31.6
11-15 years	58	38.2
16-20 years	26	17.1
Over 20 years	4	2.6
Total	152	100.0

Table 3.3 demonstrates the number of the teachers distributed to their experience in the field. The highest percentages belong to teachers whose experiences were between 6-10 years (n=48) and 11-15 years (n=58). The lowest percentage belongs to the teachers who had experience for over 20 years (n=4). Table 3.4. demonstrates the highest degree teachers hold.

Table 3.4. *Highest Degree Teachers Hold*

Highest Degree	Frequency	Percentage
Bachelor's	143	94.1
Master's	8	5.3
Doctorate	1	0.7
Total	152	100.0

According to Table 3.4., 94.1% of the teachers (n=143) hold BA degree while 5.3% of the teachers (n=8) hold Master of Arts (MA) and only one teacher hold Doctorate degree.

3.4. Instrumentation

In this study, mixed-methods research procedures were applied. For that reason, both quantitative and qualitative data collection instruments were utilized. As the quantitative tool, a Likert-scale questionnaire was used. On the other hand, the qualitative part of the study consisted of two open ended items in the questionnaire and EFL teachers' assessment documents used in lower-secondary schools.

3.4.1. Questionnaire

In the quantitative part of the study a Likert-scale questionnaire was adapted and developed by utilizing the studies of Anderson (1998), Aydoğdu (2007), Yang (2008), Çalışkan and Kaşıkçı (2010). In the composition process of the questionnaire, the

researcher consulted two field experts and some English language teachers in order to enhance its validity. Subsequently, some parts of the questionnaire were omitted and some were developed. For example there was a part which aimed to gather information about the teachers' in-class ELT practices. With the suggestions of the experts, that part was excluded and some other items related to EFL assessment were added instead.

In order to ensure the reliability of the questionnaire, it was piloted to 48 English Language Teachers working in lower-secondary schools in Çal, another district of Denizli. The data were analyzed by the SPSS 24 (Statistical Package for Social Sciences) with the aim of reliability analysis.

In literature there exist different suggestions for a good level of Cronbach's Alpha. However, in general the range of 0.70- 0.90 is conceivable. "If alpha is too high it may suggest that some items are redundant as they are testing the same question but in a different guise. A maximum alpha value of 0.90 has been recommended" (Tavakol and Denick, 2011, p. 54). According to George and Mallery (2003) the level of Cronbach Alpha reliability between 0.80 and 0.90 stands for *good* and above 0.90 is *excellent.* In addition Gliem and Gliem (2003) states that 0.80 level of reliability is well-advised. The reliability analysis of the study is demonstrated in Table 3.5.

Table 3.5. *Internal Consistency Reliability of Questionnaire*

Reliability Statistics	
Cronbach's Alpha	N of Items
.85	25

After the final version of the questionnaire was constructed, it was administered to the target population of the main study. The questionnaire consisted of 25 multiple choice items and two open-ended items. In the first part of the questionnaire an explanation of the study was presented to the participants. In order to formally ensure the teachers' willingness to participate in the study and share their documents, an agreement part was presented to the teachers following the explanations in the first part of the questionnaire. Right after, demographical information of the teachers were requested. This information proved teachers background in the sense of gender, education, experience and present teaching position.

In the second part of the study there were 25 five-point Likert-scale items demanding teachers' responses as (1-never, 2-rarely, 3-sometimes, 4-usually, 5-always). The first two items addressed teachers' general attitudes of assessment in terms of

accuracy and communicative competence. The other 23 items were composed of several traditional and alternative assessment types. At the bottom of the Likert-scale item there were two open-ended items. The first open-ended item requested teachers to write other types of assessment tools which were not included in the questionnaire but teachers used as assessment practices. The other open-ended question targeted information about teachers' four skills assessment practices in terms of type, frequency and scope of the assessment. Table 3.6 presents the items in the questionnaire.

Table 3.6. *Five-point Likert-scale Items and Open-ended Items Related to Assessment Practices*

Assessment1:I design my tests in order to assess accuracy
Assessment2: I design my tests in order to assess communicative competence
1-Multiple-choice questions (students select the answer from a set of options).   2-True/False questions (students select one of two choices, true or false).   3-Matching questions (students select the answers in one list that match the ones in the other list).   4-Fill-in-the-blank questions (students fill in a word or a phrase in a blank).   5-Wh- questions (students write content information depending on the question word)   6-Yes/No questions (students scrutinize a question or statement and construct a short response starting with Yes or No).   7-Translation questions (students translate the given words or sentence/s into the requested language).   8-Unscramble (students places the given letters or words in order to construct the requested word/s or sentence/s).   9-Informal question-answer (you ask students questions during the teaching and learning process).   10-Oral exams (you rate students with interviews).   11-Teacher-student conferences (you engage in a focused discussion with students about their work without giving marks).   12-Informal observations (you rate students' performance without pre-set criteria).   13-Formal observations (you rate students' performance with pre-set criteria).   14-Role-playing (an improvised conversation performed by students when given a situation).   15-Musical presentation (students sing songs or rhymes).   16-Presentations (students-created report/demonstration).   17-Portfolios (students' compilations of selected work with rating/reflection)   18-Creative writing (students-created poetry, short stories)   19-Journals (students' personal writing on self-chosen or assigned topics)   20-Projects (assignments given to students which involve the use of more time and resources than available during the normal class period)   21-Products (student-created graphs, tables, crafts, maps, web pages)   22-Self-assessment (students evaluate their own work)   23-Peer assessment (students evaluate other students' work)
24-Other testing and assessment tools you apply: 25- Please write about your ELT assessment practices. How and how often do you assess your students' English especially four skills?:

3.4.2. Document Analysis (Teachers' Assessment Documents)

Qualitative data of the study consisted of EFL teachers' assessment documents in addition to the open-ended items in the questionnaire. Teachers were invited to share whatever they had utilized as assessment tools in their exams; teachers shared EFL achievement exam papers they administered to the lower-secondary school students. Out of

152 EFL teachers 41 participants volunteered to share their assessment documents. 56 achievement tests were collected totally.

3.5. Data Collection Procedures

The data of the study consisted of a five-point Likert-scale questionnaire and EFL teachers' assessment documents. When the questionnaire was ready to administer, the information about the number of the lower-secondary schools and the number of the EFL teachers working in these schools were requested from the Provincial Directorate of National Education. The number of the lower-secondary schools in the central districts was 70. There were 286 EFL teachers working in these schools; while the number of the primary schools was 56 and there were 96 EFL teachers working in these schools.

At the beginning of the data collection procedure, it was intended to conduct the study in both primary and lower secondary schools to be able to incorporate all levels of young learners into the study. However, during the data collection process, it was decided not to include primary schools since the number of the voluntary teachers was not adequate to gather requisite data for the study.

In the week just after the end of the Spring Term, 2018 questionnaires were distributed to nearly all of the lower-secondary schools. Since data collection time was Workshop week of the teachers all around Turkey, all teachers were busy with occupational works carried out by the General Directorate of Teacher Training of MoNE. Workshop Program was carried out at the beginning of the Fall Term and at the end of the Spring Term annually in all educational institutions in Turkey.

The questionnaire was distributed to the voluntary EFL teachers in different schools so as to be filled in by the teachers after completing their workshop hours each day. Initially the researcher introduced herself and explained the aim of the study. With stressing the significance of their sincere responses to each item in the questionnaire, also the open-ended questions were emphasized to reflect their actual practices of four skills assessment. At the same time teachers were requested to share samples from whatever they used in their classes to assess students as the qualitative data of the study. Most teachers were not willing to share their documents. 41 teachers volunteered to share their assessment documents. The documents were only formal achievement tests which were administered after a few units were completed during and at the end of the semester. In this

study, the tests have been mentioned as *assessment documents* and *exam papers* interchangeably.

3.6. Data Analysis

In this study, the quantitative data were collected by a Likert-scale questionnaire. For the analysis, Statistical Package for Social Sciences (SPSS) version 24 was applied. With the help of descriptive statistics frequency, percentage and mean scores of the items were calculated. In order to find out if there were differences among teachers' preferences of assessment types depending on their demographical features of gender, experience and the highest degree they hold, it was determined whether the data were parametric or non-parametric by applying Kolmogorov-Smirnov and Shapiro-Wilk normality tests. For these tests, if level of significance is below *0,05* then it means the data are nonparametric. Results of these tests showed the data were non-parametric *(p<0.05)*. For this reason, as non-parametric tests Kruskal Wallis Test and Mann Whitney U Test were applied.

After that, in order to find out if there were differences between male and female teachers and between BA and MA degrees, nonparametric Mann Whitney U Test was applied. As for the teachers' experience, another non-parametric test, Kruskal Wallis Test was applied because there were more than two categories of experience. There are two hypotheses for Man Whitney U and Kruskal Wallis tests:

1.Null Hypothesis (H_0): There is no significant relation between variables.

2. Alternative Hypothesis (H_A): There is significant relation between variables.

The explanations for the hypotheses: *If p>0.05 accept (H_0); if p<0.05 accept (H_A).*

According to the results of these tests, for the items whose mean value of significance appeared to be *p<0.05,* there were significant difference between the groups (female-male; BA degree hold-MA degree hold).

The qualitative data in the last part of the questionnaire, namely open-ended items about teachers' assessment of four skills and the exam papers of EFL teachers were analyzed by using *content analysis*. "…it simply defines the process of summarizing and reporting written data – the main contents of data and their messages" (Cohen et al., 2018, p. 674). Even though content analysis is a form of qualitative data, in order to interpret and report the analysis, quantitative elements such as statistics are used (Drisko and Maschi, 2016).

In the data analysis process of this study, in order to enhance the rater-reliability, all the steps of coding and analyzing the data were carried out in company with two more coders who were also professionals in ELT field. The coding procedures were carried out in face-to-face interactions by the coders. Teachers' answers to the open-ended items of the questionnaire were firstly coded by the three coders simultaneously in terms of the basic units which were interrelated. Subsequently, all the coders compared the coded items and took joint decisions about the codes. Following this process, themes and subthemes were determined by the coders. Additionally, the 152 EFL teachers as participants were given sequential numbers and coded as *p1, p2, p3, etc.* A field expert in the university was consulted in order to check the coding, the themes and subthemes. Finally, these themes and sub-themes were interpreted by the researcher with the aim of answering the related research questions.

The other qualitative data consisted of teachers' exam papers as a supplementary material in order to verify the information gathered through questionnaire. Since the exam papers were written documents, in literature this type of analysis is called *document analysis*. "Document analysis is a systematic procedure for reviewing or evaluating documents…; it involves skimming (superficial examination), reading (thorough examination), and interpretation" (Bowen, 2009, pp. 27-32). By analyzing the exam papers, it was aimed to find out what type of items teachers used in their exams. In that sense *superficial examination* is applied in this study. This provided evidence to the information about the assessment types of teachers gathered through the questionnaire. In addition, by analyzing the exam papers it was aimed to find out whether or to what extent teachers assessed four skills of EFL students. By this way, teachers' responses to the open-ended questions in the questionnaire were compared to the information gathered by exam papers.

According to Bowen (2009) document analysis brings the elements of *content analysis and thematic analysis* together. With regard to this view, the content of the exam papers were under examination in this study. The themes were pre-determined since the research question addressed the teachers' implementations of four skills assessment. Exam papers were firstly examined by the researcher in order to detect the existence and frequency of four skills assessment both on grade basis (5th, 6th, 7th, 8th grades separately) and in total. In order to enhance the rater-reliability, the other two coders checked the exam papers subsequently. The next step was to detect the item types and the linguistic components they were aimed to be assessed in the exams on the grade basis and in total.

All the coders examined the exam papers in order to detect the item types and the related linguistic components. Finally, the field expert checked the detected item types and the related linguistic components by examining the exam papers. By this way, the analysis of the questionnaire, the content analysis of the open-ended questions and the analysis of the assessment documents complemented each other. In other words, with the help of the content analysis and document analysis a more detailed picture of the testing and assessment practices of the teachers were obtained.

CHAPTER IV: RESULTS

In this chapter, the results of the data analyses obtained from a questionnaire and EFL exam papers of EFL teachers are presented. The analyses of both quantitative and qualitative data were interpreted and explained in order to answer the following research questions:

1. What are the testing and assessment practices of EFL teachers working in state lower-secondary schools?

    1.a. How frequently do the EFL teachers prefer traditional paper-pencil tests and alternative ways of assessment?

    1.b. Are there any differences among teachers' preferences of assessment types in terms of their demographical features of gender, experience and the highest degree they hold?

    1.c. Which language skills of young EFL learners are assessed by EFL teachers in at state lower-secondary schools?

2. To what extent are the testing and assessment practices of EFL teachers consistent with the course outcomes stated by the Ministry of National Education in the English Language Teaching Program for the 5th, 6th, 7th and 8th grades?

In conformity with the data, both quantitative and qualitative data analyses procedures were followed in the study. Qualitative data consisted of written assessment documents, namely the exam papers administered to the young learners. In order to analyze the open-ended questions in the questionnaire and the exam papers content analyses technique was utilized. On the other hand, The Statistical Package for the Social Sciences (SPSS) version 24 was applied to the quantitative data obtained from the questionnaire; and the descriptive statistics were interpreted.

1. What are the testing and assessment practices of EFL teachers working in state lower-secondary schools?

When the participants were requested to share whatever they used as assessment documents, 41 out of 152 teachers shared 56 exam papers they administered during a semester. There were no other assessment documents than exam papers.

1. a. How frequently do the EFL teachers prefer traditional paper-pencil tests and alternative ways of assessment?

4.1. Descriptive Statistics of Assessment Types

In order to find out EFL teachers' assessment practices more in detail, three sub-questions were to be answered. The first sub-question aimed to find out the frequency of traditional and alternative assessment preferred by the EFL teachers. In this respect, the descriptive statistics of the assessment types are presented in Table 4.1.

Table 4.1. *Descriptive Statistics of Assessment Types (Traditional and Alternative Assessment.)*

Frequencies and Means*

Item Types	Never	Rarely	Sometimes	Usually	Always	Mean	Std. Deviation
Asessment1 accuracy	0	17	32	84	19	3.69	.83
Assessment2 communicative competence.	4	24	51	58	15	3.36	.95
1.Multiple choice	0	15	36	52	49	3.88	.97
2.True-false	0	4	22	68	58	4.18	.77
3.Matching	0	6	17	56	73	4.28	.81
4.Fill in the blanks	0	7	28	50	67	4.16	.88
5.Wh-question (open-ended)	0	16	55	45	36	3.66	.95
6.Yes/No (closed)	19	28	77	20	8	2.80	.99
7.Translation	48	46	37	17	4	2.23	1.09
8.Unscramble (words/sentences)	8	24	52	47	21	3.32	1.06
9.Informal question/answer	3	17	47	49	36	3.64	1.02
10.Oral exams	28	40	64	12	8	2.55	1.04
11.Teacher student conferences	30	33	55	22	12	2.69	1.17
12.Informal observations	10	28	58	43	13	3.13	1.02
13.Formal observations	13	27	54	41	17	3.14	1.10
14.Role playing	1	14	48	52	37	3.72	.95
15.Musical presentation	21	17	46	39	29	3.250	1.27
16.Presentations	7	20	69	40	16	3.25	.97
17.Portfolios	7	43	48	32	22	3.12	1.11
18.Creative writing	26	45	55	18	8	2.58	1.07
19.Journals	46	47	38	19	2	2.23	1.05
20.Projects	7	15	31	60	39	3.71	1.09
21.Products	11	21	43	49	28	3.40	1.15
22.Self-assessment	30	30	57	25	10	2.70	1.15
23.Peer assessment	17	36	55	37	7	2.87	1.05

* Traditional Test Mean: 3.87   (items 1-8)
*Alternative Assesment Mean: 2.98   (items 9-23)

In Table 4.1, the mean scores, frequencies and standard deviations of all the items are presented. The first two items in the questionnaire are related to accuracy and communicative competence. Assessment_1 item is "I design my tests in order to assess accuracy". Assessment_2 item is "I design my tests in order to assess communicative competence". The mean scores of them were very similar. However, teachers preferred assessing accuracy ($\bar{x}$=3.69) more frequently with a slight difference in mean scores compared to communicative competence ($\bar{x}$=3.36).

According to the statistics of item types, the items which had the highest mean scores were *matching ($\bar{x}$=4.28), true-false ($\bar{x}$=4.18)* and *fill-in-the blank* items *($\bar{x}$=4.16)* respectively with slight differences in the mean scores; multiple choice items *($\bar{x}$=3.88)* followed these items as the fourth most frequently preferred item; while the lowest mean scores belonged to *translation ($\bar{x}$=2.23), journals ($\bar{x}$=2.23), oral exams ($\bar{x}$=2.55)* and *creative writing ($\bar{x}$=2.58)*. Highest mean scores meant that teachers most frequently preferred those item types. Rationally, the lowest mean scores meant that teachers preferred those item types least frequently. On the other hand, the items which had the highest standard deviation score were *musical presentation (σ =1.27), teacher-student conferences (σ =1.17), self-assessment (σ =15)* and *product assessment (σ =1.15)* respectively. It meant that these items had the highest variance regarding teachers' responses. Additionally, the lowest standard deviation scores belonged to again *true-false (σ =0.77), matching (σ =0.81)* and *fill-in-the-blanks items (σ =0.88)* respectively. It meant that these items had the lowest variance regarding teachers' responses.

As it is demonstrated in Table 4.1, item types were categorized as *traditional assessment* and *alternative assessment*. When these two categories were compared in terms of teachers' frequency of preference, traditional assessment *($\bar{x}$=3.87)* was more frequently preferred by the teachers than alternative assessment *($\bar{x}$=2.98)*. Another interpretation about the statistics could be made comparing traditional and alternative assessment types within their own categories. The highest mean scores within traditional assessment were presented in the previous paragraphs in this chapter (matching, true-false, fill-in-the blanks). The lowest mean scores within traditional assessment belonged to *translation ($\bar{x}$=2.23)* and *yes/no items ($\bar{x}$=2.80)*. So it could be stated that teachers preferred these assessment types least frequently compared to the other traditional assessment types. The highest mean scores within alternative assessment types belonged to *role-plays ($\bar{x}$=3.72)* and *projects ($\bar{x}$=3.71)* while the lowest mean scores belonged to *journals* and *oral exams*.

It meant that teachers preferred the former most frequently while they preferred the latter least frequently.

It could be concluded from Table 4.1 that most of the teachers preferred assessing accuracy rather than communicative competence in their exams. In addition, traditional assessment type was more preferable for teachers than alternative assessment. Moreover, within the all assessment types, the mostly used item types were matching, fill-in-the blanks and true false items; while the least used item types were translation as a traditional assessment tool; journals and oral exams as alternative assessment tools.

4.2 Item Types in the Exam Papers

The 56 exam papers collected from EFL teachers were analyzed in order to find out both the item types used in the exams and the skills intended to assess in the exams. There were 12 papers of 5th grade, 13 papers of 6th grade, 16 papers of 7th grade and 15 papers of 8th grade in this study. In this section, item types which were detected in the EFL exam papers are presented based on the grades separately. Table 4.2 demonstrates the item types used in the exams of 5th grade EFL students and their frequencies within all 5th grade exam papers. Frequency referred to the *existence* of the item types rather than their amount per paper in the tables of this section.

Table 4.2. *Item Types in 5th Grade Exam Papers*

	Related Linguistic Components	Frequency	Total papers	Percentages %
Matching	Grammar- Vocabulary	12	12	100
Fill-in-the blank	Grammar-Vocabulary	12	12	100
Multiple choice	Grammar-Vocabulary-Reading	6	12	50
Wh- items	Grammar- Reading	4	12	33.3
Translation	Grammar	3	12	25
Unscrambling (word/sentence)	Grammar	2	12	16.6
Odd-one out	Grammar- Vocabulary	1	12	8.3
Restricted response essay (paragraph writing)	Writing	1	12	8.3

According to Table 4.2 in the EFL exams of 5th grade students, eight types of items were detected. The most preferred item types were *matching* and *fill-in-the blanks* items which were used in all the exams of 5th grade students. However, both *odd-one out* and *restricted response* items were used in only one exam. In half of the exam papers there

were multiple choice items. Wh- items, translation and unscrambling items were within the item types but less frequently used in 5th grade EFL exams. As for the linguistic components which were to be assessed in 5th grade exam papers, matching, fill-in-the blank and odd-one-out items were prepared to assess grammar and vocabulary. Multiple choice items were prepared to assess reading skill in addition to grammar and vocabulary. Wh- items were prepared to assess grammar and reading skill. Translation and unscrambling items were prepared to assess only grammar. Finally, restricted response essay were prepared to assess writing skill. Table 4.3 demonstrates the item types used in the 6th grade EFL exams.

Table 4.3. *Item Types in 6th Grade Exam Papers*

	Related Linguistic Components	Frequency	Total papers	Percentages %
Matching	Grammar- Vocabulary	13	13	100
Fill-in-the-blanks	Grammar- Vocabulary- Listening	13	13	100
Multiple choice	Grammar- Vocabulary- Reading	8	13	61.5
Wh- items	Grammar- Reading	5	13	38.4
True/False	Reading	3	13	23
Translation	Grammar- Vocabulary	2	13	15.3

According to Table 4.3, there were six types of items in the 6th grade exam papers. Same as the 5th grade, the most preferred item types by the EFL teachers were *matching* and *fill-in-the blanks* items in 6th grade EFL exams. Teachers used these item types in all the exams they administered. Another mostly used item type was multiple choice items. Eight out of 13 papers included multiple choice items. The other item types used in the 6th grade papers were wh-, true/false and translation items. As for the linguistic components which were to be assessed in 6th grade exam papers, they were similar to the ones in the 5th grade exam papers. For instance, matching and translation items were prepared to assess grammar and vocabulary. Fill-in-the blank items were prepared to assess listening skill in addition to grammar and vocabulary. Multiple choice items were prepared to assess reading skill in addition to grammar and vocabulary. Finally, true/ false items were prepared to assess only reading skill. Table 4.4 demonstrates the item types in 7th grade EFL exam papers.

Table 4.4. *Item Types in 7th Grade Exam Papers*

	Related Linguistic Components	Frequency	Total papers	Percentages %
Matching	Grammar- Vocabulary	16	16	100
Fill-in-the-blank	Grammar- Vocabulary	16	16	100
Multiple choice	Grammar- Vocabulary- Reading	10	16	62.5
Wh- items	Grammar- Reading	7	16	43.7
Unscrambling(word/sentence)	Grammar- Vocabulary	5	16	31.2
Translation	Grammar- Vocabulary	3	16	18.7
Yes/No	Reading	2	16	12.5

According to Table 4.4 seven types of items were used in the 7th grade EFL exam papers. Similar to 5th and 6th grades, *matching* and *fill-in-the blanks* items were used in all the 7th grade exam papers. Multiple choice items, one of the most preferred items, were available in 10 out of 16 7th grade exam papers. The other items types used in the 7th grade EFL exams were wh-, unscrambling, translation and Yes/No items. As for the linguistic components which were to be assessed in 7th grade exam papers, it is clear that they were prepared for similar purposes to the items in the 5th and 6th grade exam papers. For instance, matching, fill-in-the blank, unscrambling and translation items were prepared to assess grammar and vocabulary. Likewise, multiple choice items were prepared to assess reading skill in addition to grammar and vocabulary. Wh- items were prepared to assess grammar and reading. Finally, Yes/No items were prepared to assess reading skill. Table 4.5 demonstrates the item types used in the 8th grade EFL exams.

Table 4.5. *Item Types in 8th Grade Exam Papers*

	Related Linguistic Components	Frequency	Total papers	Percentages %
Matching	Grammar- Vocabulary	15	15	100
Fill-in-the blanks	Grammar- Vocabulary	15	15	100
Multiple choice	Grammar- Vocabulary- Reading	13	15	86.6
Wh- items	Grammar- Vocabulary- Reading	8	15	53.3
True/False	Reading	6	15	40
Yes/No items	Grammar- Reading	5	15	33.3
Restricted response (paragraph writing)	Writing	3	15	20
Error-correcting	Grammar	2	15	13.3
Odd-one-out	Vocabulary	1	15	6.6

According to Table 4.5 there were nine types of items in the 8th grade EFL exams. Not surprisingly, *matching* and *fill-in-the blanks* items were available in all the 8th grade exam papers. More frequently than in the other grade exam papers, multiple choice items were within the most preferred item types in the 8th grade exam papers. Differently from the item types in the exam papers of 5th, 6th, and 7th grades, in two 8th grade exam papers there were error correcting items. As for the linguistic components which were to be assessed in 8th grade exam papers, there were similarities to the ones in the 5th, 6th and 7th grade exam papers. For instance, matching and fill-in-the blank items were prepared to assess grammar and vocabulary. Multiple choice and wh- items were prepared to assess reading skill in addition to grammar and vocabulary. True/false items were prepared to assess only reading skill. Yes/No items were prepared to assess grammar and reading skill. Restricted response essays were prepared to assess writing skill. Error-correcting items were prepared to assess grammar. Finally odd-one-out items were prepared to assess vocabulary.

1.b. Are there any differences among teachers' preferences of assessment types depending on their demographical features of gender, experience and the highest degree they hold?

The second sub-question aimed to find out whether teachers' demographical features such as gender, duration experience and the highest degree they hold had any effect on their preferences of assessment types. Firstly, tests of normality (Kolmogorov-Smirnov and Shapiro-Wilk) were applied to the groups (male-female, highest degree they hold, experience). Each normality tests showed that the data were nonparametric. For this

reason non-parametric tests were applied to the groups. When there were two groups (male-female, BA-MA), Mann Whitney U test was utilized; when there were more than two groups (five types of duration of experience) Kruskal Wallis Test was utilized. The first analysis was made on teachers' gender. The results of Mann Whitney U Test and the item types which had significant difference *(p<0.05)* were presented in Table 4.6.

Table 4.6. *Mann Whitney U Test on Gender*

	Item Types	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
	True-false	1950.500	3078.500	-2.233	.026
Test Statistics[a]	Matching	1925.000	3053.000	-2.363	.018
	Fill-in-the blanks	1981.000	3109.000	-2.076	.038
	Informal question/answer	1917.000	7482.000	-2.285	.022

Table 4.7 demonstrates the descriptive statistics of Mann Whitney U Test based on gender.

Table 4.7. *Descriptive Statistics of Mann Whitney U Test based on Gender*

Group Statistics

Item Types	Gender	N	Mean	Std. Deviation	Std. Error Mean
True-false	Male	47	3.97	.79	.11
	Female	105	4.27	.75	.07
Matching	Male	47	4.06	.86	.12
	Female	105	4.39	.77	.07
Fill-in-the blanks	Male	47	3.91	.99	.14
	Female	105	4.27	.81	.07
Informal question/answer	Male	47	3.91	1.01	.14
	Female	105	3.52	1.01	.09

According to Table 4.7 mean scores showed that there were significant differences between male and female teachers' preferences of *true-false, matching, fill-in-the blank* and *informal question/answer sessions* items. While females preferred true-false, matching and fill-in-the blanks items more frequently than males; males preferred informal question/answer sessions more frequently than females.

Kruskal Wallis Test was applied to the groups of duration of experience. The experiences of teachers were categorized as *1-5 years, 6-10 years, 11-15 years, 16-20 years, and 20 years and over.* Test results showed that there were no significant differences *(p>0.05)* among groups. It means that teachers' duration of experience did not affect their preferences of assessment types. Table 4.8 demonstrates Kruskal Wallis Test results on duration of experience.

Table 4.8. *Kruskal Wallis Test on Duration of Experience*

		Chi-Square	df	Asymp. Sig.
	Asessment accuracy	5.047	4	.283
	Assessment communicative competence.	5.683	4	.224
	1.Multiple choice	4.695	4	.320
	2.True-false	4.476	4	.345
	3.Matching	3.867	4	.424
	4.Fill-in-the blanks	2.509	4	.643
	5.Wh-questions (open-ended)	2.685	4	.612
	6.Yes/No (closed)	3.527	4	.474
	7.Translation	4.333	4	.363
	8.Unscramble (Word/sentence)	1.145	4	.887
	9.Informal questioning	2.168	4	.705
	10.Oral exams	4.851	4	.303
Test Statistics[a,b]	11.Teacher-student conferences	1.721	4	.787
	12.Informal observations	6.310	4	.177
	13.Formal observations	4.073	4	.396
	14.Role-playing	5.447	4	.244
	15.Musical presentation	4.195	4	.380
	16.Presentations	.790	4	.940
	17.Portfolios	2.460	4	.652
	18.Creative writing	5.922	4	.205
	19.Journals	4.118	4	.390
	20.Projects	4.356	4	.360
	21.Products	2.085	4	.720
	22.Self_assessment	4.360	4	.359
	23.Peer_assessment	4.251	4	.373

Table 4.9 demonstrates Mann Whitney U Test results depending on the highest degree teachers hold (BA-MA).

Table 4. 9. *Mann Whitney U Test on the Highest Degree Teachers Hold*

		Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
Test Statistics[a]	Assessment Communicative Competence	282.000	10578.000	-2.537	.011
	Portfolios	309.500	10605.500	-2.256	.024
	Self-assessment	279.000	10575.000	-2.529	.011
	Peer assessment	333.000	10629.000	-2.067	.039

According to Table 4.9 teachers' frequencies of assessing communicative competence, using portfolios, self-assessment and peer assessment differed significantly depending on the higher degree they hold *(p<0.05).* Among teachers, there was only one teacher who hold Doctorate degree. SPSS program automatically ignored it since it was deficient as a group. Table 4.10 demonstrates the descriptive statistics of Mann Whitney U Test on the highest degree teachers hold.

Table 4.10. *Descriptive Statistics of Mann Whitney U Test on the Highest Degree Teachers Hold*

		N	Mean	Std. Deviation	Minimum	Maximum
Assessment of	Bachelor's	143	3.32	.93	1.00	5.00
Communicative	Master's	8	4.25	.88	3.00	5.00
Competence	Total	152	3.36	.95	1.00	5.00
	Bachelor's	143	2.65	1.13	1.00	5.00
Self-assessment	Master's	8	3.75	1.03	2.00	5.00
	Total	152	2.70	1.15	1.00	5.00
	Bachelor's	143	3.07	1.11	1.00	5.00
Portfolios	Master's	8	4.00	.92	3.00	5.00
	Total	152	3.12	1.11	1.00	5.00
	Bachelor's	143	2.83	1.04	1.00	5.00
Peer assessment	Master's	8	3.62	.91	2.00	5.00
	Total	152	2.87	1.05	1.00	5.00

According to Table 4.10 teachers who hold MA degree assessed communicative competence and used self-assessment, portfolios and peer assessment which are types of alternative assessment more frequently than the teachers who hold BA degree.

1.c. Which language skills of young EFL learners are assessed by EFL teachers in the classrooms?

4.3. Teachers' Responses to the Open-ended Questions (Four Skills Assessment)

At the end of the questionnaire there were two items which were composed of open-ended questions about EFL teachers' four skills assessment. The items included the questions below:

Item 24: Other assessment tools you use (please specify).

Item 25: Please add any additional comments about your ELT assessment practices in young learners' classrooms. How and how often do you assess your students' English especially language skills?

For the item 24, except one teacher, none of the teachers specified extra assessment tools they used in their classrooms. One of the teachers wrote: *"You specified all the assessment tools. Thank you"* (Open-ended item 24, Participant 56).

As for the item 25, teachers' comments on their four skills assessment practices were analyzed with the help of content analysis. Main and sub-themes were constructed by coding the data. Finally, all themes were interpreted by counting the percentages of the participants mentioned the sub-themes. The themes and sub-themes emerged from the participants' responses to the open-ended item 25 are shown in Table 4.11.

Table 4.11. *Themes and Subthemes*

Themes	Subthemes
1. Assessed Skills	1.a. Grammar & vocabulary
	1.b. Grammar & vocabulary & reading
	1.c. Grammar & vocabulary & reading & writing
2. How often?	2.a. Two times a semester
	2.b. Three times a semester
	2.c. Two or three times a semester
3. Type of four skills assessment	3.a. Informally during the lessons
	3.b. No four skills assessment at all
	3.c. Not mentioned
4. Reason for lack of four skills assessment	4.a. Time constrains
	4.b. LGS exam
	4.c. Students' negative feelings about four skills assessment
	4.d. High number of the students per classroom
	4.e. Need for a professional skills examination
	4.f. Common decision of the ELT Group at schools
	4.g. Not mentioned

The open-ended item 25 was about teachers' assessment practices in terms of what skills they assessed, how and how frequently they assessed these skills. For this reason, the first three themes were predetermined themes which the item 25 directed to the teachers. However, the forth theme was not intended by the researcher. It was one of the common topics teachers mentioned in their responses. Thus, it was determined as a theme. Teachers as participants were given sequential numbers in the analysis. They were coded as *p1, p2, p3* etc. in the tables.

Table 4.12 demonstrates the skills and linguistic components that teachers assessed.

Table 4.12. *Theme 1: Assessed Skills*

Assessed Skills	Participants	Percentage %
Grammar & vocabulary:	p2, p17, p54, p56, p65, p76, p79, p83, p85, p107 p110, p112, p115, p120, p135, p156, p163, p171	30
Grammar, vocabulary & reading:	p1, p3, p6, p12, p15, p21, p23, p28, p32, p42, p46, p58 p62, p86, p89, p94, p97, p100, p103, p122, p129, p134, p141, p150, p155 p178,	43.3
Grammar, vocabulary, reading & writing:	p4, p5, p14 p19, p38, p40, p51, p54, p69, p73, p91, p92, p138, p147, p167, 182	26.6

According to Table 4.12 teachers stated three groups of linguistic component they assessed. 30 % of the teachers *(n=18)* stated they assessed only grammar and vocabulary. For example one of the teachers stated that "*Of course I teach four skills in the lessons. But there are three exams and I assess grammar and vocabulary via them*" (Open-ended item 25, Participant 163). 43.3% of the teachers (n=26*)* stated they assessed grammar, vocabulary and reading. As an example, one of the teachers stated that "*There are two or three exams in a semester. Not all the skills, we assess just reading, grammar, vocabulary*" (Open-ended item 25, Participant 141). 26.6% of the teachers (n=16) stated they assessed reading and writing in addition to grammar and vocabulary. For example one of the teachers stated that "*I make exams two or three times. I assess reading and writing, grammar and vocabulary*" (Open-ended item 25, Participant 167). None of the teachers stated that they assessed listening and speaking in the exams. Considering the teachers' responses about the skills assessment, it could be inferred that grammar, vocabulary reading and writing were within the assessed skills. However, teachers did not assess listening and speaking in the EFL exams.

Table 4.13. *Theme 2: How Often?*

Frequency of the exams	Participants	Percentage %
Twice a semester:	p1, p2, p3, p4, p5, p6, p12, p14, p42, p83, p89, p91, p94 p100, p115, p120, p138, p150, p171, p178, 182	35
Three times a semester:	p17, p21 p38, p40, p51, p58, p97, p103, p107, p110, p122, , p129, p134, 135, p147, p155, p156, p163	30
Twice or three times a semester:	p15, p19, p23, p28, p46, p56, p69, p86, p112, p141, p167, p177	20
Not mentioned:	p32, p54, p62, p65 p73, p76, p79, p85, p92	15

According to Table 4.13 teachers stated three different frequencies of exam: two times, three times and two or three times in a semester. 35% of the teachers (n=21) acknowledged that they assessed their students' EFL twice a semester. 30% of the teachers (n=18) stated they assessed their students' EFL three times a semester. Finally, 20 % of the teachers (n=12) stated they assessed their students' EFL twice or three times a semester. 15% of the teachers (n=9) did not mention the frequency of the exams they administered. From the teachers statements it could be concluded that most teachers assessed students' English twice a semester while some teachers assessed students' English three times a semester. Table 4.14 demonstrates the types of assessment teachers specified about how they assessed their students' four skills.

Table 4.14. *Theme 3: Types of Four Skills Assessment*

Assessment Types	Participants	Percentage %
Informally during lessons:	p1, p2, p12, p17, p21, p23, p28, p40, p42, p46, p51, p54, p69, p79, p83, p89, p100, p107, p110, p115, p129, p138, p156, p163,	40
No four skills assessment at all:	p4, p5, p15, p32, p38, p54, p58, p62, p76, P91, p92, p97, p103, p112, p138, p147, p150, p155, p167, p171, p177	35
Not mentioned:	p3, p6, p14, p19, p56, p65, p73, p85, p86, p94, p97, p134, P135, p178, p182	25

According to Table 4.14, 40% of the teachers did not administer any formal exams for four skills assessment. However, they remarked that they assessed their students' four skills informally during the lessons. About this issue one of the teachers remarked that *"I observe my students in the lessons and give one or two grades according to their performance of four skills. There are three formal exams in a semester and we assess grammar and vocabulary"* (Open-ended item 25, Participant 110). Parallel to this statement one of the teachers pointed out that *"In the lessons, I assess my students' four skills. However, this is not to give mark but to contribute to their language development. In the exams I assess grammar and vocabulary* (Open-ended item 25, Participant 79).

As it is shown in Table 4.14, 35% of the teachers stated that they did not assess EFL students' four skills. For instance, about this issue, one of the teachers stated that *"I prepare exams two times in a semester. Students don't want to speak or write anything,*

they like listening mostly in the lesson. I don't assess four skills. I assess grammar and vocabulary" (Open-ended item 25, Participant 171).

25% of the teachers did not mention how they assessed EFL students' four skills. Regarding the teachers' responses to the open-ended item 25, it could be said that a considerable number of teachers preferred assessing students' four skills informally during the class periods rather than in the exams. On the other hand, some teachers clearly stated that they did not assess four skills either formally or informally. Table 4.15 presents the reasons why EFL teachers did not assess their students' four skills and their percentages as well.

Table 4.15. *Theme 4: Reasons for Lack of Four Skills Assessment*

Reasons	Participants	Percentages %
Time constrains: p14, p58, p112, p147, p150, 177		10
LGS exam: p32, p54, p56, p97, p134, p178, p182		11.6
Students' negative feelings towards four skills assessment: p42, p115, p138, p156, p171		8.3
High number of the students per class: p15, p100, p120, 135		6.6
Need for a professional skills examination: p97, p155		3.3
Common decision of the ELT group at schools: p6, p17		3.3
Not mentioned: the rest of the participants		56.9

The themes presented in Table 4.15 were detected in the EFL teachers' responses to open ended question 25 in the questionnaire. It was not intended to find out the reason of lack of four skills assessment. However, nearly half of the teachers mentioned this issue. Thus, it was presented in this study.

According to Table 4.15, 10% of the teachers stated that they did not assess four skills of EFL students because of the limited time allocated to English lessons. For instance, one of the teachers stated that *"English lesson is limited to four hours a week together with the elective English. It is impossible to teach and assess four skills in such a limited time. I assess grammar, reading and vocabulary in the three exams. I make exercises of skills in the lessons"* (Open-ended item 25, participant 58).

According to Table 4.15 another reason for lack of four skills assessment is LGS exam which is the High School Entrance Exam. 11.6% of the EFL teachers pointed out that since LGS assessed just accuracy instead of language skills; they did not assess language skills, either. One of the teachers, who commented on this issue, stated that *"I assess my students three times in a semester. The exams include vocabulary, reading and*

grammar questions. We need to prepare the students for the LGS exam and it doesn't assess four skills" (Open-ended item 25, participant 134).

8% of the teachers stated that the reason why they did not asses four skills was their students' negative feelings towards four skills assessment. About this issue, one of the teachers remarked that *"I use technological devices most of the time during the lessons for speaking and listening practices. Making a formal exam scares our students. I assess their grammar and vocabulary in the exams"* (Open-ended item 25, participant 156). Similarly one of the teachers stated that *"We assess our students two times a semester. 90 % of the exams include grammar and vocabulary in my school. We practice four skills in lessons. However, skills exams are threatening for pupils, so we give grades according to their performance in the class"* (Open-ended item 25, Participant 115).

Another reason for lack of four skills assessment is high number of the EFL students at schools. For that matter one of the teachers stated that *"I teach four skills. However, we prepare common English exams three times. It is difficult to assess four skills of 300 students in these exams. So we can assess just grammar and vocabulary"* (Open-ended item 25, participant 135). Two of the teachers (3.3%) stated that there should be a professional four skills exam administered by the Ministry of Education; that is why they did not assess four skills of EFL students. One of the teachers stated that *"I think four skills assessment requires professionalism, ministry should assess four skills. I make exams three times full of grammar, reading and vocabulary questions"* (Open-ended item 25, participant 97).

Two of the teachers (3.3%) stated that the ELT group of teachers at their school decided not to assess four skills that was the reason why they did not assess four skills. One of these teachers stated that *"The ELT teachers group in my school decides the content of the exams. We prepare two exams one of which is final exam full of multiple choice questions about vocabulary, grammar, reading; the other one is classical written exam"* (Open-ended item 25, Participant 6). Nearly half (56.9%) of the teachers did not mention any reasons for lack of four skills assessment.

Considering the teachers' comments on four skills assessment, it could be inferred that nearly half of the teachers had some reasons not to assess four skills. Some teachers thought the schedule for EFL is not enough to be able to teach and assess four skills. Some teachers had trouble about the crowded EFL classes. LGS exam was another reason since it did not involve four skills assessment. For some teachers, their students were not eager to practice four skills. It can be interpreted that the teachers did not assess these skills in the

exams in order to avoid negative backwash effect. Some teachers were of the idea that the Ministry should administer four skills exams for young learners. Lastly, the ELT groups' common decisions about not to assess four skills was another reason for EFL teachers at lower-secondary schools.

4.3. Assessment Documents of EFL Teachers (Exam Papers)

There were 56 exam papers collected from EFL teachers. The exam papers were analyzed in order to determine whether EFL teachers assessed four skills or not and detect the item types used in the exams. It was found out that there were not any exam papers which were prepared especially for four skills assessment. Namely there were not any separate exam papers for any of the four skills. All the papers included formal questions for assessment of a few skills together.

Each paper was numbered within the grades they belonged to. Before analyzing, firstly, the exam papers were grouped based on the grades they were administered to. Then, each part of the exam papers was examined in order to find out which skill or skills were assessed. If there were questions assessing one of the skills, for example, that skill was ticked in the list of skills and grades. In this study grammar and vocabulary components of language were treated as *skills* in addition to four skills; since they are also included in the assessment practices. Furthermore, since speaking skill cannot be assessed in written form together or as integrated with the other skills in the exam papers, it was not included in neither the analyses of the exam papers nor the tables which demonstrate the existence and frequency of skills assessment. Table 4.16 demonstrates the assessed skills in the exam papers based on all four grades.

Table 4.16. *Skills in the Exam Papers*

Grade& Paper no.	Listening	Reading	Writing	Grammar	Vocabulary
5th _1	-	✓	-	✓	✓
5th _2	-	-	-	✓	✓
5th _3	-	✓	-	✓	✓
5th _12	-	✓	-	✓	✓
6th _1	-	-	-	✓	✓
6th _2	-	-	-	✓	✓
6th _3	-	✓	-	✓	✓
6th _13	✓	✓	-	✓	✓
7th _1	-	-	-	✓	✓
7th _2	-	-	-	✓	✓
7th_3	-	✓	-	✓	✓
7th _16	-	✓	-	✓	✓
8th _1	-	✓	-	✓	✓
8th _2	-	✓	-	✓	✓
8th _3	-	✓	-	✓	✓
8th _15	-	✓	-	✓	✓

Since there were 56 exam papers totally, the number of them was too high to show their analysis in a table. In order to briefly demonstrate the analysis of the exam papers in terms of the skills they included, the first three and the last exam papers of each grade level were written in the table. Information in Table 4.16 was summarized in a table showing frequencies and percentages of skills which were assessed within all four grades together and in different tables showing frequencies and percentages of assessed skills for each grade separately. Table 4.17 demonstrates the frequencies and percentages of assessed skills at 5th grades. *Frequency* referred to the existence of the skills per exam paper.

Table 4.17. *Frequencies of Assessed Skills Based on 5th Grade*

Skills	Frequency within 5th Grade papers	Total papers	Percentages %
Listening	0	12	0
Reading	6	12	50
Writing	1	12	8.33
Grammar	12	12	100
Vocabulary	12	12	100

According to Table 4.17 in any of the 5th grade papers, there were not parts assigned to listening skill (0%). In half (n=6) of the 5th grade papers there were parts assigned to reading questions. In all the 5th grade exam papers (n=12) there were parts assigned to assess grammar and vocabulary. Finally, in one (8.33%) of the 5th grade exam papers there was a part in which the students were requested to write a paragraph on a given topic which was intended to assess writing skill. Table 4.18 demonstrates the frequencies of assessed skills in the 6th grade EFL exam papers.

Table 4.18. *Frequencies of Assessed Skills based on 6th Grade*

Skills	Frequency within 6th Grade papers	Total papers	Percentages %
Listening	1	13	7.69
Reading	6	13	46.1
Writing	3	13	23
Grammar	13	13	100
Vocabulary	13	13	100

According to Table 4.18, likewise the 5th grade exam papers, in the entire (100%) 6th grade exam papers there were questions prepared for both grammar and vocabulary assessment. In nearly half (46.1%) of the 6th grade papers there were parts involving questions for reading assessment. In three of the 6th grade exam papers (23%) there was a

part assigned to writing a paragraph. Surprisingly in one (7.69%) of the 6th grade exam papers there was a part which involved questions for listening assessment about an audio-record. Table 4.19 demonstrates the frequencies of assessed skills in the 7th grade EFL exam papers.

Table 4.19. *Frequencies of Assessed Skills based on 7th grade*

Skills	Frequency within 7th Grade papers	Total papers	Percentages %
Listening	0	16	0
Reading	9	16	56.2
Writing	0	16	0
Grammar	16	16	100
Vocabulary	16	16	100

According to Table 4.19, the percentages of the assessed skills at 7th grade are similar to the percentages at 5th and 6th grades. For example, in all (100%) exam papers at 7th grade there were parts which involved questions for both grammar and vocabulary assessment. Similar to 5th and 6th grade papers, there were no single questions assigned to listening and writing skills (0%) in any of the 7th grade exam papers. In nine (56.2%) of the7th grade papers there were parts assigned to assess reading skill. Table 4.20 demonstrates the frequencies of assessed skills in 8th grade EFL exam papers.

Table 4.20. *Frequencies of Assessed Skills based on 8th grade*

Skills	Frequency within 8th Grade papers	Total papers	Percentages %
Listening	0	15	0
Reading	15	15	100
Writing	3	15	20
Grammar	15	15	100
Vocabulary	15	15	100

According to Table 4.20, the results of 8th grades are very similar to the results of 5^{th,} 6th and 7th grade exam paper analyses. For example, again grammar and vocabulary were assessed in the entire (100%) 8th grade EFL exam papers. In 7th grade papers, there was a remarkable difference from the other grades in the percentage of reading assessment: The entire exam papers involved reading assessment (100%). It was exactly the same for listening skill (0%) in 8th grade EFL exam papers as in the 5th 6th and 7th grades. Namely, there were no questions assigned to listening skill assessment in the 8th grade EFL exam

part assigned to writing a paragraph. Surprisingly in one (7.69%) of the 6th grade exam papers there was a part which involved questions for listening assessment about an audio-record. Table 4.19 demonstrates the frequencies of assessed skills in the 7th grade EFL exam papers.

Table 4.19. *Frequencies of Assessed Skills based on 7th grade*

Skills	Frequency within 7th Grade papers	Total papers	Percentages %
Listening	0	16	0
Reading	9	16	56.2
Writing	0	16	0
Grammar	16	16	100
Vocabulary	16	16	100

According to Table 4.19, the percentages of the assessed skills at 7th grade are similar to the percentages at 5th and 6th grades. For example, in all (100%) exam papers at 7th grade there were parts which involved questions for both grammar and vocabulary assessment. Similar to 5th and 6th grade papers, there were no single questions assigned to listening and writing skills (0%) in any of the 7th grade exam papers. In nine (56.2%) of the7th grade papers there were parts assigned to assess reading skill. Table 4.20 demonstrates the frequencies of assessed skills in 8th grade EFL exam papers.

Table 4.20. *Frequencies of Assessed Skills based on 8th grade*

Skills	Frequency within 8th Grade papers	Total papers	Percentages %
Listening	0	15	0
Reading	15	15	100
Writing	3	15	20
Grammar	15	15	100
Vocabulary	15	15	100

According to Table 4.20, the results of 8th grades are very similar to the results of 5th, 6th and 7th grade exam paper analyses. For example, again grammar and vocabulary were assessed in the entire (100%) 8th grade EFL exam papers. In 7th grade papers, there was a remarkable difference from the other grades in the percentage of reading assessment: The entire exam papers involved reading assessment (100%). It was exactly the same for listening skill (0%) in 8th grade EFL exam papers as in the 5th 6th and 7th grades. Namely, there were no questions assigned to listening skill assessment in the 8th grade EFL exam

papers. Finally, there were three (20%) exam papers in which there was a part requesting students to write a paragraph about a given topic. Table 4.21 demonstrates the frequencies and percentages of assessed skills at all four grades.

Table 4.21. *Frequencies of Assessed Skills within All Four Grades (5th, 6th, 7th, 8th)*

Skills	Frequency	Total Paper	Percentage %
Listening	1	56	1.78
Speaking	0	56	0
Reading	36	56	64.2
Writing	7	56	12.5
Grammar	56	56	100
Vocabulary	56	56	100

According to Table 4.21, in only one (1.78%) of the 6th grade EFL exam papers out of all four grades (5th, 6th, 7th, and 8th) papers (n=56) there was a part for listening skill assessment. In none of the papers there were questions assessing speaking skill. In 36 out of 56 papers (64.2%) there were parts including questions for reading skill. In seven (12.5%) of the total papers there was a part which required students to write a paragraph about a given topic. In all of the papers (100%) there were grammar and vocabulary questions.

With regard to the percentages of assessed skills in the 56 exam papers in total, it is clear that EFL teachers tended to assess grammar and vocabulary in all exams while they did not assess listening and speaking except for one 6th grade exam paper involving a listening part. For speaking skill assessment, since teachers did not share any separate assessment document for speaking skill and it cannot be assessed in written form together or as integrated with the other skills it can be inferred that EFL teachers did not assess speaking skill at all. As for the reading skill assessment, at 8th grade it was assessed in all exam papers just like grammar and vocabulary. However, at the other grades percentage of reading assessment decreased below 50%. For writing assessment, it could be inferred that its percentage was pretty low; namely teachers did not prefer assessing writing skill regularly at lower-secondary schools.

CHAPTER V: DISCUSSION, CONCLUSION AND SUGGESTIONS

In this chapter the findings of the study will be discussed under the titles of the research questions. Subsequently, the study will be summarized and concluded. Additionally, pedagogical implications and suggestions will be provided.

5.1. Discussion

Main findings of the study are discussed providing answers to the research questions.

5.1.1. Research Question 1: What are The Testing and Assessment Practices of EFL Teachers Working in State Lower-Secondary Schools?

Both quantitative and qualitative data were analyzed and it was found out that despite the proposals of the Policy which insistently emphasize communicative language testing and alternative ways of assessment in a harmony with the other possible assessment tools, teachers utilized merely the traditional paper and pencil exams. Although the teachers asserted that they used alternative ways of assessment together with the traditional types in the questionnaires, the only assessment tools shared by the teachers were exam papers rather than materials of alternative assessment such as rubrics or scales used for portfolios, projects, products inter alia. On this basis, it could be inferred that EFL teachers working in state lower-secondary schools assessed their students by applying achievement tests in certain periods during a semester. Even though they made use of alternative types of assessment to some extent, they did not use them properly.

In literature, the number of the studies conducted on the testing and assessment of young EFL learners is limited to a few studies (Pandian, 2002; Yang, 2008; Brumen et al., 2009; Yildirim and Orsdemir, 2013; Han and Kaya, 2014; Basok, 2017). Among those studies, in Yildirim and Orsdemir's (2013) study is different from our study in terms of assessment practices since four skills assessment of young EFL learners via performance tasks was under examination in their study. In the other studies listed above, similar to our study, both traditional and alternative assessment types were implied to be practiced by the EFL teachers.

5.1.2. Sub-Question of Research Question 1: How Frequently do the EFL Teachers Prefer Traditional Paper-Pencil Tests and Alternative Ways of Assessment?

The descriptive statistics of the questionnaire revealed that the most frequently used items were matching, true/false, fill-in-the blanks and multiple choice items by the

teachers. On the other hand, the least frequently used assessment tools were translation, journals, oral exams and creative writing. It was apparent that teachers more frequently applied traditional types of assessment tools than the alternative ones. Additionally, since EFL teachers did not utilize alternative ways of assessment such as journals, oral exams and creative writing regularly, it gave us the clue that they did not assess students' writing and speaking skills sufficiently. Over and above, the frequencies of traditional and alternative assessment types were compared by grouping the items in two categories. Results supported that the mean scores of traditional assessment were higher than those of the alternatives. It meant EFL teachers mostly applied traditional pen and paper tests while assessing young learners in lower-secondary schools.

Document analyses of the teachers' exam papers were parallel to the findings of the descriptive statistics. The exam papers were analyzed in order to detect the item types used by the teachers. Accordingly, it was determined that the most frequent items were matching, fill-in-the blanks and multiple choice items. That is to say, document analysis enabled us to crosscheck the findings of the questionnaire: there was a perfect match between the findings of these data. It was remarkable that in every single exam paper there were matching and fill-in- the blank items and all of them were prepared to assess grammar and vocabulary.

With the help of document analysis, item types were compared based on the grades of students (5th, 6th, 7th, 8th). There were no differences in terms of the most frequent items. However, while the true/false items were within the most frequently preferred items according to the questionnaires, in the papers of 5th and 7th grades, there were no true/false items. What is more, multiple choice items were in the third place in terms of percentage in the exam papers of each grade level.

Yang's (2008) study had similar purposes with our study. In both researches, teachers' preferences of traditional and alternative types of assessment were under question. In Yang's (2008) study, findings indicated that matching, multiple choice, fill-in-the blank and true/false items were mostly preferred by the EFL teachers likewise the findings of our study. As for the alternative assessment tools, in both of the studies role-plays were among the most preferred tools while journals were among the least preferred tools. However, when the mean scores of traditional and alternative assessment were compared there was a difference between the studies: Yang (2008) concluded that Malaysian EFL teachers mostly applied alternative types of assessment in young learners' classrooms while our study proved the opposite.

In the study of Brumen et al. (2009), having some similarities to our study, Slovenian EFL teachers mostly preferred fill-in-the blank items; Czech EFL teachers mostly preferred true/false items while Croatian teachers mostly preferred repeat-and-drill practices. Additionally Han and Kaya's (2014) study had similar findings to our study. For example, in both studies true/false and matching items were mostly preferred traditional assessment tools in order to assess reading skill by the EFL teachers. Moreover, in both of the studies, teachers reported that role-plays were mostly preferred alternative assessment tools in order to assess speaking skill.

5.1.3. Sub-Question of Research Question 1: Are There any Differences among Teachers' Preferences of Assessment Types in Terms of Their Demographical Features of Gender, Experience and the Highest Degree They Hold?

The data revealed that there were significant differences *(p<0.05)* between male and female teachers' preferences of true/false, matching, fill-in-the blanks types of items and informal question-answer sessions. Female teachers preferred more frequently the traditional types of assessment tools, namely true/false, matching and fill-in-the blanks items than male teachers did. On the other hand, male teachers preferred informal question-answer sessions more frequently than female teachers. It gave us the clue that male teachers tended to use alternative types of assessment more than female teachers did.

Based on the experiences of teachers there were no significance differences among teachers with regard to their preferences of traditional and alternative types of assessment.

Lastly, based on the highest degrees teachers hold, the assessment types which had statistically significant difference rates were self-assessment, portfolios and peer-assessment. Teachers with MA degree preferred these alternative types of assessment tools more frequently than the teachers with BA degree. Moreover, another item that scored significantly different means in terms of teachers' educational degrees was about *communicative competence*. In that matter, teachers who hold MA degree more frequently preferred to assess communicative competence than communicative accuracy while it was the other way round for teachers who hold BA degree.

Among the studies in literature, Han and Kaya (2014) compared the findings based on gender. In their study, there were not significant differences between male and female teachers' practices of four skills assessment. In none of the studies in the related literature, the analyses were performed based on more than one demographical feature of the participants.

5.1.4. Sub-Question of Research Question 1: Which Language Skills of Young EFL Learners are Assessed by EFL Teachers at State Lower-Secondary Schools?

On the one hand, the first two items in the questionnaire were to find out whether teachers designed their assessments in order to assess accuracy or communicative competence: Teachers mostly preferred assessment for accuracy. That was the first clue about teachers' practices of assessment. On the other hand, according to the open-ended items in the questionnaire, most of the teachers (43.3%) assorted that they assessed only grammar, vocabulary and reading skills in the exams they administered to young EFL learners. When we crosscheck that data with the document analysis of exam papers, we found out that grammar and vocabulary were assessed in all grade levels with 100% of percentage. On the other hand, reading (64.2%) was in the third place within the skills according to the percentages in all grade levels. That is to say document analysis exactly supported the results of open-ended data analysis. More than that, in 8th grade exam papers percentage of reading skill was 100%. The percentage of reading skill in the exam papers increased as the grade level grew. In other words, teachers gave fewer places to reading skill in their exam papers at 5th and 6th grades compared to 7th and 8th grades. It was surprising that in only one exam paper of 6th grades there was a part allocated to listening practice. Some of the teachers (30%) acknowledged that they assessed merely grammar and vocabulary rather than four skills. The rest of the teachers stated reading and writing in addition to grammar and vocabulary were the skills they assessed in the exams. In most cases, teachers are aware of the necessity of developing communicative competence of learners. On the contrary their implementations of teaching and assessment do not reflect their pedagogical knowledge and perspectives about language teaching and learning.

All in all, teachers did not administer any separate skills exams in lower-secondary schools. They prepared exams in which grammar, vocabulary and reading had the greatest majority. Teachers did not assess speaking and listening skills of young learners at any grade levels with an exception of one exam paper including a listening part. As for the writing skill, the total percentage of it was 12.5% within all the grade levels.

Brumen et al. (2009) indicated similar findings to this study's findings. Slovenian teachers mostly assessed grammar and vocabulary and made use of fill-in-the blanks type of items in their EFL exams. Czech teachers put the emphasis on literacy skills (reading and writing) and overused true/false items in the exams of young learners; Contradictorily

Croatian teachers ignored literacy skills and focused on oral skills in company with repeat-and-drill exercises of vocabulary.

In Basok's (2017) study, there were similar findings to our study's findings. For instance, EFL teachers declared that they designed structure-based exams including grammar and reading assessment and they ignored communicative skills of listening and speaking because of the pressure by the central language examinations administered by the government and contributed to the students' success while entering a higher educational institution.

In the study reported by Pandian (2002), EFL teachers prepared exams including grammar, vocabulary, reading and writing assessment and ignored listening and speaking skills likewise in Basok's (2017) study and in our study.

Additionally, Yildirim and Orsdemir's (2013) study had similar findings to our study to some extent in terms of four skills assessment. For instance, EFL teachers ignored speaking, listening and reading skills totally while preparing performance assessment in young learners' classrooms; they just included grammar and writing in the performance tasks. In both studies listening and speaking skills were ignored. However, there is a difference: in our study reading skill was among the mostly assessed skills while in Yildirim and Orsdemir's (2013) study it was not assessed via performance tasks by the EFL teachers.

In Han and Kaya's (2014) study, the findings were very similar to the aforementioned studies and our study in terms of four skills assessment on the grounds that listening and speaking assessment were totally ignored and reading and writing skills were assessed through the exams prepared by the EFL teachers working with young learners.

5.1.5. To What Extent are the Testing and Assessment Practices of EFL Teachers Consistent with the Course Outcomes Stated by the Ministry of National Education in the English Language Teaching Program for the 5th, 6th, 7th and 8th grades?

In their responses to the open-ended items in the questionnaire teachers commented on how and how frequently they assessed the skills they preferred to assess at all grade levels of lower-secondary schools. Their comments were analyzed with the help of content analysis. There occurred four main themes with sub-themes. The first theme was related to the skills they assessed during a semester. It was interpreted while answering the third sub-question of the first question. Most of the teachers (43.3%) asserted that they assessed merely reading as a skill; additionally, grammar and vocabulary were assessed extensively

as the linguistic components. The oral skills listening and speaking were completely ignored. Together with the results of the quantitative data it could be inferred that EFL teachers tended to implement grammar-based traditional paper and pencil assessment procedures. In parallel with our study, Basok (2017) examined the consistency of the curriculum to the implementations of EFL teachers working in primary, secondary and high schools. Teachers ingeniously declared that their assessment implementations did not match the policy. They designed grammar-based assessment practices since they felt under the pressure of high-stake language examinations. Likewise in Yildirim and Ordemir's (2013) study, teachers tended to assess grammar and writing rather than all four skills in their implementations of performance tasks. Their study indicated a mismatch between the curriculum proposals and the teachers working with young learners. On the contrary, in the ELT policy of the MoNE (2018), communicative competence has emphasis. For assessment, MoNE (2018, pp. 6-7) suggests all kinds of testing and assessment techniques such as summative and formative, product and process-oriented tests and traditional and alternative assessment tools which cover four skills and other linguistic components by taking students' cognitive developments and motivational issues into consideration.

Another main theme in teachers' assertions was the frequency of their four skills assessments. Most of the teachers (35%) stated that they prepared exams twice a semester while the rest of the teachers stated as *three times* and *twice or three times* in a semester. Needless to say, aforementioned exams did not cover all four skills. MoNE fixed the number of the achievement exams as two in a semester for all primary and lower-secondary schools in January, 2018. Considering that this study was conducted at the end of the spring semester of 2018, it would be expected that all teachers declared the number of the exams as twice a semester.

In their responses, teachers commented on how they treated the skills they did not include in the exams. Most teachers (40%) asserted that they assessed especially speaking and listening skills informally during the lessons. On the other hand, the percentage (35%) of the teachers who acknowledged the inexistence of four skills in assessment practices at all was noteworthy. The contradiction between the in-class practices and assessment procedures put forth a trouble in the validity and reliability of the exams. Even though teachers attempted to integrate the language skills into their teaching, assessment practices lacking those skills caused a mismatch even between their own practices of teaching and assessment before the policy. In none of the studies in the related literature, teachers

mentioned their way of treating the skills which they ignored in the exams. For this reason, it could be inferred that our study obtained different findings from the other studies.

Within the assertions of EFL teachers in the open-ended items, an extra theme was worth noting. It was not intended to gather such a data by the researcher since research questions of the study did not include such a question. However, it embodied significant clues about why four skills assessment could not be implemented as suggested by the Ministry. Approximately half of the EFL teachers declared some excuses for the lack of a proper four skills assessment in young learners' classrooms they taught EFL. Actually teachers complained about some contradictories in the EFL education system. Teachers pointed out that schedule of the EFL lessons was not enough; the classrooms were overcrowded; students had negative feelings about skills assessment especially speaking skill; both teachers and students felt under pressure by the LGS exam which included structure-based EFL questions; teachers needed ready-made four skills tests supplied by the MoNE; and finally the ELT groups at lower-secondary schools did not include four skills assessment into their agenda. These were voices of teachers adding remarkable details to the scene. Even though not relevant to the aims of the study, they were presented with the hope that they would be heard by authorities or whoever to take actions to make things better.

All in all, the implementation of the assessment proposals of the MoNE in the young learners' classrooms proved to be problematic in this study. To provide a whole answer to the final question of this study, teachers tended to use mostly traditional paper and pencil tests based on grammar and vocabulary rather than a harmony of all kinds of assessment techniques and tools to foster communicative competence of the young learners. In that sense, it is apparent that a huge gap exists between the course outcomes of ELT Program stated by the MoNE and teachers' practices of EFL assessment in young learners' classrooms.

In Basok's (2017) study, teachers declared that the underlying reason for the mismatch between the policy and the EFL assessment practices of teachers was that central language examinations only assess grammar and reading and ignore the communicative skills. In that sense, our study produced similar findings with Basok's (2017) study since some of the EFL teachers in our study indicated LGS exam as the underlying reason for not having communicative skills in their assessment practices. Similar to our study and Basok's (2017) study, in the study reported by Pandian (2002), teachers stated that they design structure-based tests since listening and speaking did not take place in the Primary

Schools Assessment Test administered at the end of the 6th grade. In that case, it could be concluded that high-stake examinations stand for one of the reasons behind the mismatch between the policy and practices of EFL assessment in young learners' EFL education.

5.2. Conclusion

CEFR has been accepted and implemented as a pathfinder in our country since 2006. ELT Program has been revised and innovated several times in terms of language teaching learning and assessment (e.g. 1997, 2005, 2012, 2015 and 2018). Considering the policy innovations in the ELT Program of MoNE, this study aimed to find out the assessment practices of EFL teachers in lower-secondary schools (5th, 6th, 7th, and 8th grades) by examining what types of assessment tools EFL teachers use in young learners classrooms and whether there are differences in the preferred assessment tools depending on teachers' demographical features such as gender, experience and higher degree they hold. Another aim of the study was to find out the extent of consistency between proposed course outcomes of the ELT Program stated by MoNE and the EFL teachers' assessment practices in lower-secondary schools. The research questions of the study were as follows:

1. What are the testing and assessment practices of EFL teachers working in state lower-secondary schools?

1.a How frequently do the EFL teachers prefer traditional paper-pencil tests and alternative ways of assessment?

1.b Are there any differences among teachers' preferences of testing and assessment tools in terms of their demographical features of gender, experience and the highest degree they hold?

1.c Which language skills of young EFL learners are assessed by EFL teachers at state lower-secondary schools?

2. To what extent are the testing and assessment practices of EFL teachers consistent with the course outcomes stated by the Ministry of National Education in the English Language Teaching Program for the 5th, 6th, 7th and 8th grades?

The sample of the study was composed of 152 EFL teachers working in state lower-secondary schools in the central districts of Denizli. The design of the study was parallel databases design under the convergent parallel approach in mixed research methods. The quantitative data were collected by a five-point Likert-scale questionnaire including open

ended items. As the qualitative data, the open-ended items in the questionnaires were responded by 60 out of 152 EFL teachers. Furthermore, teachers were requested to share all kinds of assessment documents they used in classrooms as another qualitative data, and 41 out of 152 teachers shared 56 exam papers of 5th, 6th, 7th, and 8th grades.

Questionnaires were administered to 152 EFL teachers working in state lower-secondary schools. Descriptive analysis of the questionnaires as the quantitative data was performed by the SPSS version 24 program. The open ended items in the questionnaires and finally the exam papers as document analysis were analyzed by making use of content analysis.

With the help of the questionnaires, information about the EFL teachers' practices of EFL assessment was gathered. Through the Likert-scale type of items in the questionnaire, the frequencies of traditional and alternative assessment tools utilized by the EFL teachers were found out. Through non-parametric Mann Whitney U Test it was found out whether there was a significant difference between male and female teachers' preferences of traditional and alternative assessment tools. In addition, whether there was a significant difference between teachers' preferences of traditional and alternative assessment tools depending on the highest degree teachers hold (BA and MA degrees) was found out by applying Mann Whitney U Test. On the other hand, through another non-parametric test, Kruskal Wallis Test, it was found out whether there were significant differences among teachers in terms of using traditional and alternative assessment tools depending on the duration of their experiences. On the other hand, through the open-ended items in the questionnaires, the language skills assessed by the EFL teachers in young learners' classrooms were found out.

As for the exam papers, firstly the item types used in the exams were detected with the help of two other coders and a field expert in the university. Subsequently, four main themes with several subthemes related to four skills assessment practices of the teachers were determined. Through the interpretations of the themes, the language skills assessed in the exams, the frequency of the exams and how teachers administered four skills assessment in young learners classrooms were found out. Additionally, some teachers stated the reasons for not including all four skills in the EFL assessment in young learners' classrooms although it was not intended to find out by the researcher.

Findings of the descriptive statistics revealed that EFL teachers used traditional assessment more frequently than alternative assessment. Furthermore, female teachers tended to use traditional assessment more frequently than male teachers did. In the same

way, male teachers tended to use alternative assessment more frequently than female teachers did. On the other hand, teachers with MA degree preferred alternative assessment more frequently than the teachers with BA degree. There were no significant differences in teachers' assessment preferences of traditional and alternative assessment depending the duration of their experience. As for the item types, teachers mostly preferred matching, true/false, fill-in-the blank and multiple choice items respectively according to the mean scores of the items. The least preferred items were translation, journals, oral exams and creative writing. Moreover, the findings of the descriptive statistics indicated that teachers preferred to assess accuracy more frequently than communicative competence.

According to the results of content analysis of the open-ended items, the EFL teachers assessed grammar and vocabulary in all the achievement exams they administered to the young learners. In addition to grammar and writing, reading and writing skills were assessed by the EFL teachers to some extent. Listening and speaking skills were not included in the assessment practices of the EFL teachers. Some teachers declared that they assessed listening and speaking skills informally during the lessons. Furthermore, the findings of the document analysis substantially supported these findings. Namely, in the exam papers, the skills assessed with a hundred percent by the teachers were grammar and vocabulary. In nearly half of the exam papers, there were parts for reading assessment. Additionally, in some exam papers writing skill was assessed. However, in none of the exam papers there were questions for listening assessment. As for the speaking skill, since it cannot be assessed through written materials, and none of the teachers declared they assessed speaking skill, it was interpreted that teachers did not assess speaking skill at all.

Above and all, the findings of this study revealed that the assessment practices of the EFL teachers working in state lower-secondary schools did not match the CEFR-oriented ELT policy of the MoNE since the EFL teachers tended to design traditional structure-based tests instead of a harmony of all kinds of assessment tools and techniques based on communicative competence.

5.3. Pedagogical Implications

Turkey has passed through a number of innovations in the educational policy of English Language Teaching. Adopting a CEFR-oriented policy has brought remarkable responsibilities to the stakeholders especially EFL teachers. Additionally, the recent revisions in the curriculum, has provided a well-organized model for teachers. Thus the implementations of the proposed teaching and assessment procedures have become under

question. However, it stands for a new issue especially in testing and assessment of what teachers actually do in the classrooms. This study is one of the few studies examining the assessment practices of EFL teachers of young learners and their consistency to the policy in Turkey such as Yildirim and Orsdemir's (2013), Han and Kaya's (2014), and Basok's (2017) studies. For this reason, this study might contribute to comprehend the policy and implement it in a more compatible way in Turkish context.

The findings of the study may provide feedback to teachers, teacher trainers, and policy makers in order to find a common ground. MoNE may provide EFL teachers with in-service training in terms of communicative competence, communicative language assessment, characteristics of young learners and integration of four skills into both teaching and assessment processes. Model testing and assessment practices may be presented by the MoNE. Additionally testing and assessment of four skills may become compulsory for primary and lower-secondary schools. Teachers may take actions to improve their pedagogical knowledge and skills through MA programs or any other helpful courses.

5.4. Suggestions

Above all, it is significant to determine the underlying reasons of that inconsistency between policy and practice to be able to produce applicable solutions. Therefore this study gives way to further research on investigating the background problems of this inconsistency. Researchers may conduct more researches focusing on the underlying reasons of such problems in the implementation of the policies at schools.

By modeling this study, researchers may compare their findings to this study in order to provide a wider picture of the situation in Turkey. They may conduct similar studies aiming to determine the consistency or inconsistency by making some changes in the methodology of this study. For example, the data of this study was collected through questionnaires and exam papers of teachers; they may collect the data through interviews and observations in the real classroom environments. They may also extend the number of the participants as well as the type of the schools such as primary and high schools. After conducting a similar study, a model for four skills testing and assessment may be attached to the study after applying it in EFL classrooms. Since there are few studies examining the assessment practices of EFL teachers working with young learners in the international contexts such as the studies of Yang (2008), Pandian (2002) and Brumen et al. (2009), this study provides a model for the researchers in the international context in order to see the

situation in their own countries. Needless to say, it is necessary to conduct further research to be able to elaborate on the findings of this study.

REFERENCES

Alderson, J.C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Arslan, R. Ş. (2012). Bridging the gap between policy and practice in teaching English to young learners: The Turkish context. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi,* 32(32), 95–100.

Anderson, R. S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New Directions for Teaching and Learning*, 1998(74), 5–16. https://doi.org/10.1002/tl.7401

Arikan, A. (2015). The CEFR and reading : A document analysis. *Procedia - Social and Behavioral Sciences*, 199, 501–504. https://doi.org/10.1016/j.sbspro.2015.07.538

Brumen, M., Cagran, B., & Rixon, S. (2009). Comparative assessment of young learners' foreign language competence in three Eastern European countries. *Educational Studies,* 35(3), 269–295. https://doi.org/10.1080/03055690802648531

Bachman, L., & Palmer, A. (2010). *Language assessment in practice.* Oxford: Oxford University Press.

Benigno, V., & De Jong, J. H. (2016). The global scale of English learning objectives for young learners: A CEFR- based inventory of descriptors. *Educational Linguistics*, 2016(25), 43-64 https://doi.org/10.1007/978-3-319-22422-0

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. https://doi.org/10.3316/QRJ0902027

Brown, H.D. (2004). *Language assessment. Principles and classroom practices*. NY, USA:Pearson Education.

Brown, H.D. (2007). *Teaching by principles. An interactive approach to language pedagogy.* NY, USA: Pearson Education.

Burns, A. & Siegel, J. (2018). *International perspectives on teaching the four skills in ELT: Listening, speaking, reading, writing*. Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1177/003368827600700219

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.

Carless, D. R. (2003). Factors in the implementation of task-based teaching in primary schools. *System*, *31*(4), 485–500. https://doi.org/10.1016/j.system.2003.03.002

Cheng, L., & Fox, J. (2017). *Assessment in the language classroom. Teachers supporting student learning.* UK: Macmillan Education, *6*(2), 200.

Clark, J.L.D. (1972). *Foreign language testing: theory and practice*. Philadelphia: Center for Curriculum Development.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). London and New York: Routledge.

Council of Europe (CoE). (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, England: Cambridge University Press.

Council of Europe (CoE). (2018). *Common European framework of reference for languages: Learning, teaching, assessment*. *Companion Volume with New Descriptors.* Cambridge, England: Cambridge University Press.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative and mixed methods approaches.* Thousand Oaks, CA: SAGE Publications.

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches. *SAGE Publications, Incorporated.* (p. 265). https://doi.org/10.1002/macp.200400177

Crystal, D. (1997). *English as a global language*. Cambridge: Cambridge University Press.

Çalişkan, H., & Kaşikçi, Y. (2010). The application of traditional and alternative assessment and evaluation tools by teachers in social studies. *Procedia - Social and Behavioral Sciences*, 2(2), 4152–4156. https://doi.org/10.1016/j.sbspro.2010.03.656

Damar, E. A. (2004). *A study on teaching english to young learners in EFL contexts*. Unpublished M.A.Thesis, Uludağ University Bursa, Turkey.

Damar, E., Gursoy, E., & Korkmaz, S. (2013). Teaching English to young learners: Through the eyes of teacher trainer. *ELT Research Journal*, *2*(3), 95–110.

Doğançay-Aktuna, S., & Kiziltepe, Z. (2005). English in Turkey. *World Englishes*, *24*(2), 253–265. https://doi.org/10.1111/j.1467-971X.2005.00408.x

Drisko, J., & Maschi, T. (2016). *Content analysis*. Oxford: New York: Oxford University Press.

Edmonds, W, A., & D. Kennedy, T. (2017). *An applied guide to research designs: Quantitative, qualitative, and mixed methods* (Second Edition). Los Angeles: SAGE Publications.

East, M. (2008). *Dictionary use in foreign language writing exams: impact and implications.* Amsterdam, Netherlands/Philadelphia, PA: John Benjamins.

Fulcher, G. & Davidson, F. (2007). *Language Testing and Assessment: An Advance Book* (Vol. 7). https://doi.org/10.1177/026553228500200106

Fulcher, G., & Davidson, F. (2007). The Common European Framework of Reference (CEFR) and the design of language tests : A matter of effect. *Language Teaching*, 40, pp 231-241 doi:10.1017/S0261444807004351

Fulcher, G. (2010). *Practical language testing*. London: Hodder Education.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference.* 11.0 update (4th ed.). Boston: Allyn & Bacon.

Gliem, J. A., Gliem, R. R., & Gliem, J. A. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for likert-type scales, *2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education*, (1992), 82–88. https://doi.org/10.1109/PROC.1975.9792

Halliwell, S. (1992). *Teaching English in the primary classroom*. London: Longman.

Han, T., & Kaya, H. İ. (2014). Turkish EFL teachers' assessment preferences and practices in the context of constructivist instruction. *Journal of Studies in Education*, *4*(1), 77. https://doi.org/10.5296/jse.v4i1.4873

Harmer, J. (2007). *The practice of English language teaching*. Essex: Pearson Education.

Hasselgren, A. (2005). Assessing the language of young learners. *Language Testing*, *22*(3), 337–354. https://doi.org/10.1191/0265532205lt312oa

Heaton, J.B. (Eds.) (1988). *Writing English language tests* (Second Edition). New York: Longman.

Henning, G. (1987). *A Guide to language testing: development, evaluation and research*. Rowley, Massachusetts: Newbury House.

Hymes, D. H. (1966). *Two types of linguistic relativity.* In Bright, W. *Sociolinguistics*. The Hague, The Netherlands: Mouton & Co. pp. 114–158.

Ionnoau-Georgiou, S. & Pavlou, P. (2003). *Assessing young learners*. Oxford: Oxford University Press.

Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education*, 6(2), 26–37. Retrieved from http://www.eab.org.trhttp//ijrte.eab.org.tr

Kırkgöz, Y. (2007). English language teaching in Turkey: policy changes and their implementations. *RELC Journal*, 38(2), 216-228.

McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.

Ministry of National Education. (2013). İlköğretim İngilizce dersi (2, 3, 4, 5, 6, 7 ve 8.sınıflar) öğretim program [Primary education English language curriculum (the 2nd-8th grades)]. Retrieved from http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=327.

Ministry of National Educcation. (2015).İlköğretim İngilizce Dersi Öğretim Programı (İlkokul ve Ortaokul 2, 3, 4, 5, 6, 7 ve 8. Sınıflar). Ankara: MEB Yayıncılık

Ministry of National Education. (2018). İlköğretim İngilizce dersi (2, 3, 4, 5, 6, 7 ve 8. sınıflar) öğretim program [Primary education English language curriculum (the 2nd-8th grades)]. Retrieved from http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=327.

Ministry of National Education. (2018). Öğretmen Yetiştirme Genel Müdürlüğü Öğretmenlerin 2018 Eylül Dönemi Mesleki Çalişma Programi.[General Directorate of Teacher Training and Educatio] Retrieved from http://oygm.meb.gov.tr/www/2015-2018-mesleki-calisma-programi/icerik/693

Ministry of National Education - Board of Education. (2017). İlköğretim kurumları haftalık ders çizelgesi [Weekly course schedule of primary education institutions (the 2nd - 8th grades)]. Retrieved from http://ttkb.meb.gov.tr/www/haftalik-derscizelgeleri/kategori/7

Nikolov, M. (2016). *Assessing young learners of English: Global and local perspectives*. Switzerland: Springer International Publishing DOI10.1007/978-3-319-22422-0

Oller, J.W. (1976). Language testing. In R. Wardhaugh and H.D. Brown (Eds.), *A survey of applied linguistics* (275-300). Ann Arbor: University of Michigan Press.

Paker, T. (2012). Türkiye'de neden yabancı dil (İngilizce) öğretemiyoruz ve neden öğrencilerimiz iletişim kurabilecek düzeyde İngilizce öğrenemiyor? [Why do not we teach foreign language (English), and why cannot our students learn English at a certain level to communicate?]. *Pamukkale University Journal of Education*, 32(2), 89-94.

Paker, T. (2013). The backwash effect of the test items in the achievement exams in preparatory classes. *Procedia - Social and Behavioral Sciences*, *70*, 1463–1471. https://doi.org/10.1016/j.sbspro.2013.01.212

Pandian, A. (2002). English language Teaching in Malaysia today. *Asia Pacific Journal of Education*, 22:2, 35-52, DOI: 10.1080/0218879020220205

Promodou, L. (1995). The backwash effect: From testing to teaching. *English Language Teaching Journal*, 49:1, 13–25.

Rea-Dickins, P. (2000). Assessment in early years language learning contexts. *Language Testing* 17, 115–22.

Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and wellbeing. *American Psychologist*, 55(1), 68–78.

Sarıgöz, İ. H., & Fişne, F. N. (2018). English language assessment and evaluation practices in the 4th grade classes at mainstream schools. *Journal of Language and Linguistic Studies*, 14(3), 380-395.

Shaaban, K. (2005). Assessment of young learners. *English Teaching Forum*, *43*(1), 34–40. Retrieved from http://americanenglish.state.gov/resources/english-teaching-forum-2005-volume-43-number-1#child-567

Shohamy, E., Or, I. G., & May, S. (2017). *Language Testing and Assessment*. (Third Edition). Switzerland: Springer International Publishing.

Simonson M., Smaldino, S, Albright, M. and Zvacek, S. (2000). Assessment for distance education (ch 11). *Teaching and Learning at a Distance: Foundations of Distance Education.* Upper Saddle River, NJ: Prentice-Hall.

Slattery, M., & Willis, J. (2001). *English for primary teachers.* Oxford: Oxford University Press.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Trim, J. (2011). *Using the CEFR : Principles of Good Practice*. Cambridge: Cambridge ESOL

Tsagari, D. (2004). Is there life beyond language testing ? An introduction to alternative language assessment. *Centre for research in language education. CRILE,* Working Paper No 58.

Underhill, N. (1987). *Testing Spoken Language: A Handbook of Oral Ttesting Techniques.* Cambridge: Cambridge University Press.

Vale, D. and Feunteun, A. (1995). *Teaching children English*. Cambridge: Cambridge University Press.

Vygotsky, L. S. (1978). *Thought and language*. (Rev. and edited by A. Kozulin). Cambridge, MA: MIT Press.

Yazıcı, M. (2014). Değerler ve toplumsal yapida sosyal değerlerin yeri / Values and position of social values in social structure. *Firat Üniversitesi Sosyal Bilimler Dergisi Journal of Social Sciences*. 24. 209-223.

Yildirim, R., & Orsdemir, E. (2013). Performance tasks as alternative assessment for young EFL learners: Does practice match the curriculum proposal? *International Online Journal of Educational Sciences*, *5*(3), 562–574. Retrieved from http://ezproxy.scu.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=93436242&site=ehost-live

APPENDICES

APPENDIX A: EFL Testing & Assessment Questionnaire

Dear Colleague,

This survey is intended to elicit your practices of testing and assessment in English as a Foreign Language (EFL) courses . The information you provide is considered confidential and will be used for research purposes only. You do not need to mention your name. We truly appreciate your volunteering to cooperate and spend time completing the questionnaire.   *Meral ÜÇOK ATASOY*

Please put a tick and confirm your permission for the information and documents you voluntarily share to be used in this research, if you agree to the explanations .

Yes, I agree ☐

Background Information

Please mark the item that best describes your background and current teaching situation. (Check all items that apply.)

Gender:

Male ☐   Female ☐

Present Teaching Position:

Primary school (2nd-3rd-4th grades) ☐   Secondary school (5th -6th-7th-8th grades) ☐

Teaching Experience:

1-5 years ☐   6-10 years ☐   11-15 years ☐   16-20 years ☐   Over 20 years ☐

Education:

☐ Faculty of Education, Department of ELT

☐ Open Education Faculty, Department of ELT

☐ Faculty of Science and Literature

- ✓ Department of English Language and Literature ▢
- ✓ Department of American Culture ▢
- ✓ Department of English Linguistics ▢
- ✓ Other (please specify) …………………………………………………………..

☐Other teaching branches (Mathematics teaching, Turkish teaching, Primary School Teaching, Science Teaching etc.) Please specify ………………………………………………….

☐Another department which is not directly related to English or teaching (Physics, Chemistry, Engineering, Business etc.) Please specify……………………………………………

The highest degree you have obtained:

Bachelor's ☐   Master's ☐   Doctorate ☐

**If you would like to receive a report of this survey, please provide your e-mail address:

(…………………………………………..@..)

SURVEY

Part 1

Following are various types of classroom assessment. Please respond to each item by using the following response scale. Circle the number that best describes your own assessment practices with young learners.

	I apply the following assessment types:	Always	Usually	Sometimes	Rarely	Never
*	I design my tests in order to assess accuracy	5	4	3	2	1
*	I design my tests in order to assess communicative competence	5	4	3	2	1
1	**Multiple-choice questions** (students select the answer from a set of options).	5	4	3	2	1
2	**True/False questions** (students select one of two choices, true or false).	5	4	3	2	1
3	**Matching questions** (students select the answers in one list that match the ones in the other list).	5	4	3	2	1
4	**Fill-in-the-blank questions** (students fill in a word or a phrase in a blank).	5	4	3	2	1
5	**Wh- questions** (students write content information depending on the question word)	5	4	3	2	1
6	**Yes/No questions** (students scrutinize a question or statement and construct a short response starting with Yes or No).	5	4	3	2	1
7	**Translation questions (**students translate the given words or sentence/s into the requested language)	5	4	3	2	1
8	**Unscramble** (students places the given letters or words in order to construct the requested word/s or sentence/s).	5	4	3	2	1
9	**Informal question-answer sessions** (you ask students questions during the teaching and learning process).	5	4	3	2	1
10	**Oral exams** (you rate students with interviews).					
11	**Teacher-student conferences** (you engage in a focused discussion with students about their work without giving marks).	5	4	3	2	1
12	**Informal observations** (you rate students' performance without pre-set criteria).	5	4	3	2	1
13	**Formal observations** (you rate students' performance with pre-set criteria).	5	4	3	2	1
14	**Role-playing** (an improvised conversation performed by students when given a situation).	5	4	3	2	1
15	**Musical presentation** (students sing songs or rhymes).	5	4	3	2	1
16	**Presentations** (students-created report/demonstration).	5	4	3	2	1
17	**Portfolios** (students' compilations of selected work with	5	4	3	2	1

	rating/reflection)					
18	**Creative writing** (students-created poetry, short stories)	5	4	3	2	1
19	**Journals** (students' personal writing on self-chosen or assigned topics)	5	4	3	2	1
20	**Projects** (assignments given to students which involve the use of more time and resources than available during the normal class period)	5	4	3	2	1
21	**Products** (student-created graphs, tables, crafts, maps, web pages)	5	4	3	2	1
22	**Self-assessment** (students evaluate their own work)	5	4	3	2	1
23	**Peer assessment** (students evaluate other students' work)	5	4	3	2	1

Part 2

Please provide responses to the open-ended items below.

24	**Other testing and assessment tools you apply in the classrooms:**
25	**Please write about your ELT assessment practices. How and how often do you assess your students' English especially four skills?:**

End of the survey. Thank you very much! ☺

PERSONAL INFORMATION

Name	Meral
Surname	Üçok Atasoy
Birth place/date	Amasya 30/05/1989
Nationality	T.C.
Contact address and e-mail address	Sırakapılar Mah. 1531. Sok. No:10/4 Merkezefendi-DENİZLİ meralucok@gmail.com
Education	
Primary	Mehmet Bilgili İlköğretim Okulu-Suluova İlköğretim Okulu (1995-2003)
Secondary	Sinop Anadolu Öğretmen Lisesi- Amasya Anadolu Öğretmen Lisesi (2003-2007)
Higher education (Bachelor's degree)	Orta Doğu Teknik Üniversitesi (2007-2011)
Higher education (Master's degree)	Pamukkale Üniversitesi, Yabancı Diller Eğitimi ABD (2014-)
Foreign Language	
Foreign language	English
Exam name	YDS
Exam date	September, 2014
Points received	96,25
Professional Experience	
2011-2012	Beykent University
2012-2017	Denizli/Çal Denizler İlkokulu/ Ortaokulu
2017-still	Denizli/Merkezefendi Necla-Ergun Abalıoğlu Mesleki ve Teknik Anadolu Lisesi