



T.C

BURSA ULUDAĞ UNIVERSITY

INSTITUTE OF EDUCATIONAL SCIENCES

FOREIGN LANGUAGES TEACHING DEPARTMENT

ENGLISH LANGUAGE TEACHING

**A COMPARATIVE ANALYSIS OF TEACHERS' BELIEFS AND
PRACTICES ON THE ASSESSMENT OF 4TH GRADE-EFL STUDENTS
IN TURKEY, ITALY AND FINLAND**

MASTER'S THESIS

Nilay ÇAKIR

BURSA

2020



T.C

BURSA ULUDAG UNIVERSITY

INSTITUTE OF EDUCATIONAL SCIENCES

FOREIGN LANGUAGES TEACHING DEPARTMENT

ENGLISH LANGUAGE TEACHING

**A COMPARATIVE ANALYSIS OF TEACHERS' BELIEFS AND
PRACTICES ON THE ASSESSMENT OF 4TH GRADE-EFL STUDENTS
IN TURKEY, ITALY AND FINLAND**

MASTER'S THESIS

Nilay ÇAKIR

Thesis Advisor

Prof. Dr. Zübeyde Sinem GENÇ

BURSA

2020

BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim.



Nilay ÇAKIR

05/02/2020



EĞİTİM BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS İNTİHAL YAZILIM RAPORU

ULUDAĞ ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

YABANCI DİLLER EĞİTİMİ ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 02/03/2020

Tez Başlığı / Konusu: Türkiye, İtalya ve Finlandiya'daki Öğretmenlerin, 4. Sınıf-EFL Öğrencilerinin Değerlendirmelerine Yönelik Görüş ve Uygulamalarının Karşılaştırmalı Analizi

Yukarıda başlığı gösterilen tez çalışmamın a) Giriş, b) Ana bölümler ve c) Sonuç kısımlarından oluşan toplam 128 sayfalık kısmına ilişkin, 02/03/2020 tarihinde şahsım tarafından (Turnitin)* adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan özgünlük raporuna göre, tezimin benzerlik oranı % 12'dir.

Uygulanan filtrelemeler:

1- Kaynakça hariç

2- Alıntılar hariç

3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Özgünlük Raporu
Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu uygulama esaslarında
belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini;
aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul
ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Adı Soyadı : Nilay ÇAKIR 02 / 03 / 2020 Öğrenci nu : 801510008

Anabilim Dalı : Yabancı Diller Eğitimi

Anabilim Dalı Programı : İngilizce Öğretmenliği

Programı Statüsü : Yüksek Lisans



Prof. Dr. Zübeyde Sinem GENÇ

03/03/2020

YÖNERGEYE UYGUNLUK ONAYI

“Türkiye, İtalya ve Finlandiya’deki Öğretmenlerin, 4. Sınıf-EFL Öğrencilerinin Değerlendirmelerine Yönelik Görüş ve Uygulamalarının Karşılaştırmalı Analizi” adlı Yüksek Lisans tezi, Bursa Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırlanmıştır.



Tezi Hazırlayan

Nilay ÇAKIR



Danışman

Prof. Dr. Zübeyde Sinem GENÇ



Yabancı Diller Eğitimi Anabilim Dalı Başkanı

Prof. Dr. Zübeyde Sinem GENÇ

T.C.

BURSA ULUDAĞ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE,

Yabancı Diller Eğitimi Anabilim Dalı'nda 801510008 numara ile kayıtlı Nilay ÇAKIR'ın hazırladığı "Türkiye, İtalya ve Finlandiya'daki Öğretmenlerin, 4. Sınıf-EFL Öğrencilerinin Değerlendirmelerine Yönelik Görüş ve Uygulamalarının Karşılaştırmalı Analizi" konulu Yüksek Lisans çalışması ile ilgili tez savunma sınavı, 08/05/2020 gün 13.00-14.00 saatleri arasında yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin / çalışmasının **başarılı** olduğuna **oybirliği** ile karar verilmiştir.

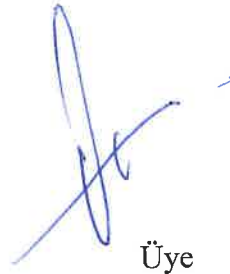


Üye (Tez Danışmanı ve Sınav Komisyonu Başkanı)

Prof. Dr. Zübeyde Sinem GENÇ



Dr Öğr. Üyesi Kürşat CESUR



Dr Öğr. Üyesi Derya YILMAZ

Önsöz

Türkiye'de yapılan bazı çalışmalar, yetkililer tarafından resmi olarak önerilen bir program olmasına rağmen öğretmenlerin, öğrenci değerlendirmesinde hala kağıt-kalem testleri gibi geleneksel metodları kullandığını, öz-değerlendirme ya da portfolyo gibi yeni metodların gözardı edildiğini göstermiştir. Ayrıca, sınavların içeriği kelime ve dil bilgisine dayalı olup, dinleme ve konuşma becerilerini ihmal etmektedir. Bu çalışmalara benzer olarak, dünyanın farklı yerlerinde yapılan araştırmalarda da genç öğrencilerin yabancı dil yeterliliklerinin hala sayısal ölçeklerle geleneksel yöntemler kullanılarak değerlendirildiği belirtilmiştir. Bu çalışmanın amacı Türkiye, İtalya ve Finlandiya'daki ilkökul yabancı dil öğretmenlerinin, yabancı dil değerlendirme sistemiyle ilgili düşünceleri ve uygulamalarını belirlemek ve yapılan uygulamaların teorik olarak hedeflenen şekilde olup olmadığını ortaya çıkarmaktır.

Öncelikle tez çalışmamda planlanmasında, araştırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteğini esirgemeyen, engin bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren danışman hocam Prof. Dr. Zübeyde Sinem Genç'e teşekkürlerimi ve saygılarımı sunarım.

Tez çalışmam sürecinde benden desteğini bir an için bile esirgemeyen sevgili eşime ve tüm eğitim hayatım boyunca maddi ve manevi olarak her zaman yanımda olan değerli abime teşekkür ederim.

Son olarak bu çalışmada kullanılan enstrümanların uygulanması sürecinde etkin rol alan katılımcı öğretmen arkadaşlarıma minnetlerimi sunarım.

Nilay Çakır

Özet

Tezin Adı : Türkiye, İtalya ve Finlandiya'daki Öğretmenlerin, 4. Sınıf-EFL Öğrencilerinin Değerlendirmelerine Yönelik Görüş ve Uygulamalarının Karşılaştırmalı Analizi

Ana Bilim Dalı : Yabancı Diller Ana Bilim Dalı

Üniversite Adı : Bursa Uludağ Üniversitesi

Enstitü Adı : Eğitim Bilimleri Enstitüsü

Tez Kabul Tarihi :

Danışmanı : Prof. Dr. Zübeyde Sinem GENÇ

TÜRKİYE, İTALYA VE FİNLANDIYA'DAKİ ÖĞRETMENLERİN, 4. SINIF-EFL ÖĞRENCİLERİNİN DEĞERLENDİRMELERİNE YÖNELİK GÖRÜŞ VE UYGULAMALARININ KARŞILAŞTIRMALI ANALİZİ

Türk eğitim sistemindeki yeni değişikliklerle birlikte, İngiliz dili öğretimi müfredatı öğrencilerin ihtiyaçları göz önüne alınarak tekrar düzenlenmiştir. Yeni müfredat daha çok iletişimsel yeterliliklere odaklanmakta ve öğrenci port folyosu, öz-değerlendirme, akran değerlendirmesi gibi ölçme ve değerlendirme anlamında yeni metot ve teknikler içermektedir. Buna rağmen, Türkiye'de ve dünyanın farklı yerlerinde yapılan araştırmalarda öğrenci değerlendirmesinde hala sınavların içeriğinin kelime ve dil bilgisine dayalı olduğu, dinleme ve konuşma becerilerinin ihmal edildiği belirtilmektedir. Bu çalışmanın amacı Türkiye, Finlandiya ve İtalya'daki yabancı dil öğretmenlerinin, ilkokul EFL öğrencilerinin değerlendirmelerine yönelik görüş ve uygulamalarını belirlemek ve yapılan uygulamaların teorik olarak hedeflenen şekilde olup olmadığını ortaya çıkarmaktır. Araştırmanın örneklemini Türkiye'den seçilen toplam 89 İngilizce öğretmeni, İtalya'dan seçilen toplam 40 İngilizce öğretmeni ve Finlandiya'dan seçilen toplam 41 İngilizce öğretmeni oluşturmuştur.

Veri toplama aracı olarak, iki bölümden oluşan anket ile açık uçlu on sorudan oluşan yarı yapılandırılmış görüşme formu kullanılmıştır. Görüşme yoluyla elde edilen nitel veriler içerik analizine tabi tutulurken, anket yoluyla toplanan nicel veriler frekans, yüzde, ortalama, standart sapma ve tek yönlü ANOVA kullanılarak analiz edilmiştir.

Araştırmanın sonuçları, öğretmenlerin değerlendirme uygulamalarının müfredat ve kitaplar, tutumlar, sınıf mevcudu, ders süresi, öğrenci özellikleri ve var olan materyaller gibi değişkenlere bağlı olarak değiştiğini göstermiştir. Farklı ülkelerde sınıftan sınıfa değişim gösterse de, öğretmenlerin genç öğrencileri değerlendirmedeki zorlukları temel olarak öğrenme ortamından, öğretim teknolojisi ve materyallerden, öğrencilerden, ebeveyn algısı, ders saati ve değerlendirme aracından kaynaklanmaktadır. Bunun yanı sıra, görüşmelerin analizlerine göre, öğretmenler ders kitaplarından alınan ve kullanıma hazır olan sınavları ve sınavlara yönelik ek materyalleri kullanma eğilimindedir ve öğretmenlerin çoğunluğu, biçimlendirici değerlendirme türlerinin kullanılmasının öğrenciler için daha uygun olduğunu ifade etmişlerdir.

Anahtar sözcükler: Değerlendirme, EFL Öğretmen Tutum ve Uygulamaları, Finlandiya, İtalya, Türkiye

Abstract

Name of Thesis : A Comparative Analysis of Teachers' Beliefs and Practices on the Assessment of 4th Grade-EFL Students in Turkey, Italy and Finland

Department : Foreign Language Teaching

Name of University : Bursa Uludag University

Name of Institute : Institute of Educational Sciences

Degree Date :

Supervisor : Prof. Dr. Zübeyde Sinem GENÇ

A COMPARATIVE ANALYSIS OF TEACHERS' BELIEFS AND PRACTICES ON THE ASSESSMENT OF 4TH GRADE-EFL STUDENTS IN TURKEY, ITALY AND FINLAND

With the recent changes in Turkish educational system, English language teaching program has been redesigned taking the learners' needs into consideration. The new curriculum focuses on communicative skills more and includes some methods such as student portfolio, self-assessment and peer-assessment for evaluation. However, in the studies from Turkey and different parts of the world, the results indicate that in the assessment of young learners' foreign language competence, the content of tests is still based on grammar and vocabulary items and most of the assessments ignore listening and speaking skills. The aim of this study is to find out what foreign language teachers' beliefs and practices are with regard to the assessment of primary school EFL students and discover whether there are matches or mismatches between the expected quality of assessment and the reality in Turkey, Finland and Italy. A total of 89 English teachers in Turkey, 40 English teachers in Italy and 41 English teachers in Finland constituted the sample of the study. For data collection, a questionnaire consisting of two main parts and a semi-structured interview with ten open-ended questions

were used. The qualitative data obtained through interviews were subjected to content analysis whereas the quantitative data collected through the questionnaire were analyzed by using frequency, percentage, mean, standard deviation and one-way ANOVA.

The results of the study indicated that the use of assessment practices of teachers varied across countries according to the curriculum and the book they followed, their beliefs, classroom size, students' characteristics and the materials. Despite its variety from classroom to classroom in different countries, the difficulties teachers face in assessing young learners were mainly stemmed from learning environment, instructional technology and material, students, parent perception, class hours and assessment tools. Also, according to the analysis of the interviews, teachers tended to use ready-to-use exams taken from the course books as well as additional materials for the exams and most of the teachers suggested that using formative types of assessment were more appropriate for young learners.

Key words: Assessment, EFL Teachers' Beliefs and Practices, Finland, Italy, Turkey

Table of Contents

	Page
BİLİMSEL ETİĞE UYGUNLUK SAYFASI	i
İNTİHAL YAZILIM RAPORU SAYFASI.....	ii
YÖNERGEYE UYGUNLUK SAYFASI	iv
JÜRİ ÜYELERİNİN İMZA SAYFASI	v
ÖNSÖZ	vi
ÖZET	vii
ABSTRACT	ix
TABLE OF CONTENTS	xi
LIST OF TABLES	xvi
LIST OF FIGURES	xviii
LIST OF ABBREVIATIONS	xix
CHAPTER 1: INTRODUCTION	1
1.1. Background of the Study	1
1.2. Statement of the Problem	2
1.3. Purpose of the Study.....	4
1.4. Significance of the Study.....	4
1.5. Research Questions	5
1.6. Definition of Key Terminology.....	5
CHAPTER 2: LITERATURE REVIEW	6
2.1. Assessment	6
2.1.1. Basic concepts associated with assessment	8
2.1.1.1. <i>Formative assessment</i>	8
2.1.1.2. <i>Summative assessment</i>	9
2.1.1.3. <i>Traditional assessment</i>	10

2.1.1.4. <i>Alternative assessment</i>	10
2.1.1.5. <i>Criterion-referenced test versus norm-referenced test</i>	13
2.1.2. Qualities of instruments used for the assessment	15
2.1.2.1. <i>Reliability</i>	15
2.1.2.2. <i>Factors influencing reliability</i>	16
2.1.2.3. <i>Validity</i>	17
2.1.2.4. <i>Factors influencing validity</i>	19
2.1.2.5. <i>Practicality</i>	19
2.1.2.6. <i>Washback</i>	20
2.2. Assessment in EFL Curriculum and Classes.....	20
2.2.1. Teachers' beliefs and practices of assessment.....	23
2.3. Assessing Young Learners in ELT	28
2.4. Assessment Practices in Elementary Schools.....	32
2.4.1. EFL and assessment practices in elementary schools in Turkey	32
2.4.2. EFL and assessment practices in elementary schools in Italy and Finland.....	34
2.5. Research on Young Learners in Foreign Language Classes	38
2.5.1. Assessment practices in foreign language classroom in Turkey	38
2.5.2. Research on assessment practices in foreign language classroom in Italy	41
2.5.3. Research on assessment practices in foreign language classroom in Finland	42
2.5.4 Research on assessment practices in foreign language classroom in other countries	42
2.5.5. Research on teachers' perceptions of and beliefs about assessment	44

CHAPTER 3: METHOD	46
3.1. Research Design	47
3.2. Participants	49
3.3. Data Collection Instruments	51
3.3.1. Questionnaire.....	51
3.3.2. Semi-structured interview.....	53
3.4. Data Collection Procedures	54
3.5. Data Analysis.....	54
CHAPTER 4: RESULTS	56
4.1. Questionnaire results	56
4.1.1. Typical class size	56
4.1.2. The reasons for carrying out the assessment	57
4.1.3. Responsible body for the preparation of the assessment.....	58
4.1.4. Reporting the exam results	59
4.1.5. Additional training for the assessment in primary schools.....	60
4.1.6. Types of assessment used in FL classroom	61
4.1.7. Content / skills – focus in assessment of students	66
4.1.8. Kinds of exercises used in FL classroom	69
4.1.9. Beliefs about the types of assessment.....	74
4.1.10. Beliefs about content / skill – focus in assessment.....	79
4.1.11. Beliefs about the kinds of exercises	82
4.1.12. Teachers’ general beliefs about assessment of students’ foreign language.....	87
4.2. Interview Results	93
4.2.1. Exams used for young learners.....	94

4.2.1.1. <i>Turkish teachers</i>	94
4.2.1.2. <i>Italian Teachers</i>	95
4.2.1.3. <i>Finnish teachers</i>	95
4.2.2. The ways of assessing communicative skills.	96
4.2.2.1. <i>Turkish teachers</i>	96
4.2.2.2. <i>Italian Teachers</i>	97
4.2.2.3. <i>Finnish teachers</i>	97
4.2.3. Using exams that assess communicative skills.....	98
4.2.3.1. <i>Turkish teachers</i>	98
4.2.3.2. <i>Italian Teachers</i>	99
4.2.3.3. <i>Finnish teachers</i>	99
4.2.4. Using language portfolio as an assessment tool	100
4.2.4.1. <i>Turkish teachers</i>	100
4.2.4.2. <i>Italian teachers</i>	101
4.2.4.3. <i>Finnish teachers</i>	101
4.2.5. Thoughts on self and peer-evaluation.....	101
4.2.5.1. <i>Turkish teachers</i>	102
4.2.5.2. <i>Italian Teachers</i>	103
4.2.5.3. <i>Finnish teachers</i>	103
4.2.6. Difficulties in assessing young learners.	103
4.2.6.1. <i>Turkish teachers</i>	103
4.2.6.2. <i>Italian Teachers</i>	104
4.2.6.3. <i>Finnish teachers</i>	104
4.2.7. Solutions for the difficulties in assessing young learners.....	104
4.2.7.1. <i>Turkish teachers</i>	104

4.2.7.2. <i>Italian Teachers</i>	105
4.2.7.3. <i>Finnish teachers</i>	105
4.2.8. Types of assessment in case of proper conditions	105
4.2.9. The ways to improve assessment skills	106
4.2.9.1. <i>Turkish teachers</i>	106
4.2.9.2. <i>Italian Teachers</i>	107
4.2.9.3. <i>Finnish teachers</i>	107
4.2.10. General suggestions for assessment of young learners.....	107
CHAPTER 5: DISCUSSIONS AND CONCLUSION	109
5.1. Discussion.....	109
5.2. Summary.....	118
5.3. Limitations of the Study	122
5.4. Implications and Suggestions for Further Studies.....	122
REFERENCES.....	125
APPENDICES.....	136
ÖZGEÇMİŞ.....	146

List of Tables

<i>Table</i>		<i>Page</i>
1.	Demographic Characteristics of the Sample	50
2.	Typical Class Sizes in Three Countries	56
3.	The Reasons for Assessment	57
4.	Responsible Body for the Preparation of Assessment.....	58
5.	The Ways of Reporting Exam Results	59
6.	The Types of Education They Received for Assessment in Primary Schools.	60
7.	Frequency of Types of Assessment Used in FL Classroom (Turkey).....	63
8.	Frequency of Types of Assessment Used in FL Classroom (Italy).....	64
9.	Frequency of Types of Assessment Used in FL Classroom (Finland).....	65
10.	ANOVA Results in Terms of the Types of Assessment Teachers Used.....	66
11.	Frequency of Content / Skills – Focus in Assessment of Students (Turkey) ...	67
12.	Frequency of Content / Skills – Focus in Assessment of Students (Italy)	68
13.	Frequency of Content / Skills – Focus in Assessment of Students (Finland)...	68
14.	ANOVA Results in Terms of Types of Content / Skills Teachers Assessed at 4 th Grade Language Exams.....	69
15.	Frequency of Kinds of Exercises Used in FL Classroom (Turkey)	70
16.	Frequency of Kinds of Exercises Used in FL Classroom (Italy).....	71
17.	Frequency of Kinds of Exercises Used in FL Classroom (Finland).....	72
18.	ANOVA Results In Terms of Kinds Of Exercises Used In FL Classroom.....	74
19.	Beliefs about the Types of Assessment (Turkey).....	75
20.	Beliefs about the Types of Assessment (Italy)	76
21.	Beliefs about the Types of Assessment (Finland)	78

22.	ANOVA Results in Terms of Teachers Beliefs about Types of Assessment To Be Used in FL Classroom	79
23.	Beliefs about Content / Skill – Focus in Assessment (Turkey)	80
24.	Beliefs about Content / Skill – Focus in Assessment (Italy)	81
25.	Beliefs about Content / Skill – Focus in Assessment (Finland)	81
26.	ANOVA Results in Terms of Teachers’ Beliefs about the Content / Skills – Focus Needed To Be Assessed for 4 th Graders.....	82
27.	Beliefs about the Kinds of Exercises (Turkey).....	84
28.	Beliefs about the Kinds of Exercises (Italy).....	85
29.	Beliefs about the Kinds of Exercises (Finland).....	86
30.	ANOVA Results in Terms of Teachers’ Beliefs about Kinds of Exercises Needed To Be Used for 4 th Grade Students.....	87
31.	General Beliefs about Assessment of Students’ Foreign Language (Turkey). 90	
32.	General Beliefs about Assessment of Students’ Foreign Language (Italy).....	93
33.	General Beliefs about Assessment of Students’ Foreign Language (Finland) 92	
34.	ANOVA Results in Terms of Teachers’ General Beliefs about Assessment of 4 th Grade Students’ Foreign Language	93
35.	Types of Assessment Teachers Would Use If They Had Proper Conditions. 105	

List of Figures

<i>Figure</i>		<i>Page</i>
1.	Comparison of Alternative Assessment and Traditional Assessment.....	13
2.	Stages of Classroom-based Assessment	21
3.	Scope and Time of Assessment	23
4.	Assessment for and of Learning	25
5.	Characteristics of Young and Older Learners	30
6.	Skills and Suggested Activities by Grades in 2018 English Language Curriculum in Turkey	33
7.	Common Reference Levels	35
8.	Types of Assessment Suggested in CEFR.....	36
9.	The Steps Followed for Revision and Adaptation of the Instrument	51

List of Abbreviations

- CEFR : The Common European Framework of Reference for Languages
- EFL : English as a Foreign Language
- ELT : English Language Teaching
- ESL : English as a Second Language
- EU : European Union
- FA : Formative Assessment
- FL : Foreign Language
- GPA : The Grade Point Average
- SPSS : Statistical Package for the Social Sciences
- TESOL : Teaching English to Speakers of Other Languages
- TPR : Total Physical Response

Chapter 1

Introduction

In this chapter, the background of the study, the statement of the study, the purpose of the study, the significance of the study and also the research questions addressed are explained briefly and clearly.

1.1. Background of the Study

Assessment and evaluation have always received attention in ELT especially for the purpose of language policy, methodology and curriculum design in language teaching, language learning and also in language teachers' professional development. Most recently, assessment as a part of curriculum of all subjects has received much greater attention by language teachers and curriculum specialists due to its potential to affect the ways of instruction and evaluation of language learner (Dochy, 2001; Remesal, 2011). Assessment has a critical role and multidimensional functions in terms of enabling teachers to know their learners, determine the learning objectives, design the learning environment, select the appropriate materials and provide feedback to learners.

Assessment is a systematic way of collecting, reviewing and using a lot of information regarding to education program in order to improve learning (Palomba & Balta, 1999). In addition to collection of systematic and related information, Anandan (2015) indicates that various tools and techniques as a part of assessment help educators to document, measure and evaluate learners' readiness, progress and skills. Assessment is being used for many purposes in educational settings and procedures. Wiesnerová (2013) grouped the functions of assessment as informative, corrective and motivational. Liljedahl (2010) asserts the main purposes of assessment as communication, valuing what is taught, reporting out and not sorting / not ranking. As understood from the definition and its area of usage, assessment provides information and feedback not only for the students' progress and development in

various areas (e.g. knowledge, skills, beliefs, values etc.), but also for curriculum and course implementation (e.g. book, plans, activities, methods etc.) (Anandan, 2015; Ioannou – Georgiou, 2003; Nasab, 2015; Palomba & Balta, 1999).

In language classes, assessment is an integral part of language teaching and learning due to the fact that the information collected through assessment gives idea about the quality of instruction and progress of students in terms of having adequate knowledge and skills in language (Doğandere, 2006). Assessment in English language classes is generally used for educational purposes such as placement, guidance, prediction, diagnosis, selection, evaluation, or other administrative purposes (Lamprianou & Athanasou, 2009). Thus, assessment is used for many purposes and to take different decisions either in process, or product or in both in language classes. In this sense, teachers have critical roles for making assessment process effective and useful. Cumming (2009) indicates that as an assessor, teachers' role in assessment is rather complex and includes several components as monitoring and reporting students' achievement, diagnosing students' learning difficulties, and making student groups for a specific task or placement. In order to be affective in assessment to observe as much information as possible regarding teaching and students' progress, teachers make use of variety of assessment tools in language classes which help them observe the strengths and weaknesses of teaching and learning process, and also provide evidences for re-designing their instructions.

1.2. Statement of the Problem

EFL education at primary level has gained importance in Turkey as it has all around the world. With the recent changes in the Turkish educational system (4+4+4), English lesson has become mandatory at 2nd, 3rd and 4th grades at primary schools and English language teaching program has been redesigned taking the learners' needs into consideration. The functions which are based on communicative competence like telling, asking, responding,

naming etc. are prepared according to the A1 (basic user) level of CEFR at 2nd, 3rd and 4th grades (MEB, 2018). The new curriculum also includes some methods about evaluation such as student portfolio, self-assessment and peer-assessment. Some studies in Turkey show that although there is a program suggested by authorities for schools, teachers still use traditional methods like pen and paper tests to assess their students; self-assessment is not applied at schools and the portfolio assessment is still being neglected by the teachers (Kırkgöz & Ağçam, 2012; Öz, 2014; Tatlı, 2014). Moreover, the content of tests is based on grammar and vocabulary items and most of the assessment procedures ignore listening and speaking skills. In the studies from different parts of the world, the results indicate that in the assessment of young learners' foreign language competence, traditional written–oral assessment by numerical grades is still carried out.

Even though teachers are designing and planning their teaching processes in line with the curricula, variety of factors influence their choice of assessment type and procedures. Teachers' beliefs regarding language teaching and learning, their experience in the teaching profession, students' characteristics and also the school type were previously reported as the determinants of teachers' selection or choice of assessment (Broadfoot & Black, 2004; Pajares, 1992). Another important issue regarding the selection of assessment tools and ways is associated with teachers' conceptions and attitudes. Teachers' attitudes toward assessment influence the ways in which the assessment happens (Griffiths, Gore & Ladwig, 2006). Also, teachers' perceptions and beliefs in relation to assessment practices have a considerable effect on the ways they use in teaching and their assessment practices as well as teachers' behaviors and students' performance (Jia, Eslami & Burlbaw, 2006; Johnson, 1992).

Teachers' perspective of assessment, the context in which they implement the assessment and curriculum / syllabus followed are the critical factors which are influential on

the selection and implementation of assessment. How this is different from classroom to classroom in different countries is still in question.

1.3. Purpose of the Study

The aim of this study is to find out what foreign language teachers' beliefs and practices are with regard to the assessment of primary learners of foreign languages and discover whether there are matches or mismatches between the expected quality of assessment and the reality in Turkey, Finland and Italy. The research focused on 4th grade level because formal assessment starts at 4th grades in Turkey. Various aspects of the assessment system (types, activities, content, skills etc.) were analyzed in detail and comparatively in 3 countries to find out the problems and deficiencies that teachers have.

1.4. Significance of the Study

Turkey is in the process of being a member of the European Union and makes adaptation and reform attempts in all areas of education to be in line with EU criteria. Designing the language curricula according to CEFR is one of the steps taken for this adaptation attempts. European countries such as Italy and Finland has already been implementing CEFR criteria in language education for a long time. Therefore, comparing teachers' practices and their beliefs regarding assessment in Turkish EFL classes and European EFL classes help to understand how assessment in language classes is selected and implemented in line with CEFR criteria and how Turkish EFL teachers' practices of and beliefs about assessment differ from European teachers. This understanding of assessment practices sheds light also on classroom practices and their reflection on the teaching processes. Furthermore, this study indicates perspective about teachers' influence on their use of assessment, selection and implementation of content / skills to be acquired and also language exercises.

1.5. Research Questions

Based upon the general aim of the study, following research questions were addressed throughout the study.

1. How is the assessment being implemented by EFL teachers at 4th grades in primary schools in Turkey, Finland and Italy?
2. What are the assessment activities used by EFL teachers in Turkey, Finland and Italy?
3. What are the content and related skills in assessment in Turkey, Finland and Italy?
4. What are the teachers' beliefs about assessment of 4th grade students in Turkey, Finland and Italy?
5. What are the problems in the assessment practices in Turkey, Finland and Italy?
6. What are the teachers' deficiencies in the assessment of young learners in Turkey, Finland and Italy?

1.6. Definition of Key Terminology

Assessment: Assessment is the systematic collection, review and use of several data / information on any educational program utilized in order to improve development and learning (Palomba & Banta, 1999).

Attitudes: Attitude is the fundamental orientation to evaluate people, other living beings, things, events, and ideas along a good - bad dimension (Banaji & Heiphetz, 2010 , p.377)

Young learners: Young learners are the ones studying in primary schools in most countries and defines young learners as the children between the ages of five and twelve (McKay, 2006).

Chapter 2

Literature Review

In order to establish a strong base and understand the findings of the present study better, the concept of assessment and associated issues are presented in this chapter.

2.1. Assessment

Assessment, as one of the main components of curriculum, has a critical role to implement and monitor the learning and teaching processes effectively. Assessment has multidirectional functions such as enabling teachers to know their learners, determining the learning objectives, designing the learning environment, selecting appropriate materials and providing feedback to learners about their progress.

Assessment and evaluation have always received attention in ELT especially for language policy, methodology and curriculum design in language teaching, in language learning and also in language teachers' professional development. Performing assessment and evaluation has been considered as the responsibility of specialists. However, the paradigm shift experienced in the understanding of assessment and evaluation in the field of ELT in the last ten and fifteen years due to the changes in expectations of governments and funding agencies in all around the world has affected learners, teachers and classroom practices (Davison & Cummins, 2007). This shift seen in the field of ELT has influenced the reconceptualization of assessment in language classes. The developments and dynamic changes in language teaching due to its theoretical changes since the beginning of the twentieth century (Richards & Rodgers, 2014) have evolved the concept of assessment in language teaching and learning. Evolution of assessment has been observed in four processes (Heaton, 1990) as :

- pre-scientific stage (focusing on memorization of words and structures),

- psychometric - structuralist approach (focusing on acquisition of sets of habits and mastery of skills),
- psycholinguistic-sociolinguistic approach (focusing on using skills in an integrated way and meaningful presentation of language in a context) and
- communicative approach (focusing on using language in a real context).

Assessment and evaluation are sometimes raise confusion and used interchangeably although they are both different in meaning and used for different purposes. Some researchers have tried to differentiate these terms to avoid misunderstanding. Assessment is more related to receiving information and evidence whereas evaluation is more related to making judgment through the use of this information and / or evidence (Anandan, 2015). Ioannou – Georgiou (2003) makes a distinction between assessment and evaluation. She explains that assessment includes a variety of ways to collect information about the learners' progress (e.g. knowledge, skills etc.) whereas evaluation is a process to obtain information to judge the extent to which the course of study meets its goals. Several efforts have been dedicated to define the term of assessment due to its comprehensiveness.

Palomba and Banta (1999) view assessment as a systematic collection, review and use of data / information on any educational program utilized in order to improve development and learning.

Anandan (2015) relates assessment to the process of obtaining information systematically as part of evaluation and asserts that assessment includes various tools and techniques used by educators for documenting, measuring and evaluating readiness, progress and skills of learners in the whole educational process (from pre-school to adulthood).

Nasab (2015) perceives assessment as the gap between learning and teaching, and defines it as informal information gathering process on students' knowledge through the use of several ways to collect information in different time and contexts. Wiesnerová (2013)

mentions that assessment has a variety of functions and divides these functions into three categories as informative, corrective (regulative, prognostic and developmental) and motivational functions. Liljedahl (2010) claims that assessment serves four purposes in general as follows; communication, valuing what is taught, reporting out and not sorting / not ranking.

Based on the definitions of assessment above, it could be defined as a collection of information through the use of tests on the learners, and processing (review, use, analyze etc.) this information. Assessment does not only refer to the product, but also to the process and includes many ways to obtain information regarding learners' ability, understanding, knowledge and motivation (Ioannou – Georgiou, 2003). Thus, assessment provides information not only about students' knowledge and capacity on a certain topic but also about their tendencies, beliefs, values and so on. The information and evidence emerged from the assessment provide feedback to students to be aware of their progress and attainments and also to teachers to examine their instructions (in terms of appropriateness of setting, instructional design, method and materials) and to monitor learners.

2.1.1. Basic concepts associated with assessment. There are various concepts associated with the terms of assessment. Some concepts are related to its nature while some are related to its types.

2.1.1.1. Formative assessment. In language classes, ultimate desire of the teachers establishes a fruitful learning environment for learners to attain language skills. Although teachers make several plans and designs for the classroom, they may sometimes not realize the weaknesses in the implementation and not observe how the learners progress in acquiring language skills in an intended way. In this regard, formative assessment, as a part of the whole assessment procedures, encourages teachers to monitor the progress of learning during the process of instruction. Its purpose is to provide continuous feedback not only to teachers

but also learners concerning learning successes and failures, and also ongoing information on instructional process while implementing and learning are occurring (Anandan, 2015).

Formative assessment focuses mainly upon identifying any areas needed to be improved and goes hand in hand with the instruction. Formative assessment, as a part of learning process (Brookhart, 2003), is “*evaluating students in the process of “forming” their competencies and skills with the goal of helping them to continue that growth process*” (Brown & Abeywickrama, 2012, p.7).

Bhola, (1990) explains this procedure as a method of judging the worth of a program during its implementation and it focuses on the progress of learners, programs, courses etc. This type of assessment is not for grading, rather for identifying students’ progress and determining the effectiveness of teaching. Nasab (2015) claims that formative assessment is an integral part of the classroom instruction, and provides immediate feedback and evidence to motivate learners to attain the learning objectives.

Formative purpose of the assessment is performed in the process as well as at the end of the course of the study in order to guide the students for their own learning progress and also the teachers to modify and re-design the learning environment, teaching materials and method to fit the course objectives to the students’ needs and capabilities (Bachman & Palmer, 1996).

2.1.1.2. Summative assessment. Summative and formative types of assessment are seen to be contrasting one with the other. The main purposes of this procedure are grading, the certification of pupil achievement and also providing information for the appropriateness of the course objectives and effectiveness of the instructions. Bhola (1990) states that summative evaluation is a method of judging the value of a program at the end of the program activities. *Outcome* is the key of this procedure. Summative assessment takes place to provide feedback and evidence in order to sum up the learning and teaching progress (Anandan,

2015). Summative assessment is generally undertaken at the end of the semester / term primarily for grading the students and secondarily for presenting the feedback on the achievement of the students (Nasab, 2015).

Summative purpose of assessment is implemented for administrative purposes at the end of the school year in order to provide information about the students' progress and attainments at the end of the course of the study (Bachman & Palmer, 1996) and to assess students for the certification and provide promotion for the next level (Genesee & Upshur, 1996).

2.1.1.3. Traditional assessment. Traditional assessment is more teacher-dominated and mainly used for summative purposes. This type of assessment is inauthentic, indirect, standardized and norm-referenced (Bailey, 1998). It measures students' performance in a short period of time and thus it is called as single –occasion tests (Law & Eckes, 1995). The evidence obtained through the use of traditional assessment is somehow limited to adequate feedback on teaching and learning process, and also far beyond monitoring the students' progress in the process. It is individualized and not context specific.

Traditional assessment including true-false items, multiple choice items and so forth generally measures learners' lower order thinking and lower level learning such as memorization and recall of cognitive level of learning (Dikli, 2003).

2.1.1.4. Alternative assessment. Traditional assessment may sometimes not produce adequate information for monitoring students' progress and re-designing the learning process. In such contexts, complementary assessment procedures are needed to provide holistic understanding and perspective regarding the things being assessed. Alternative assessment including all methods and techniques associated with performance assessment, holistic assessment and authentic assessment (Lee, Park & Choi, 2011) is proposed due to the fact that the available traditional methods and techniques are not sufficient to determine learners' real

development and strengths (Balliro, 1993). Alternative assessment emerges as a response to inadequacies and shortcomings of traditional types of assessment (Al-Mahrooqi & Denman, 2018) and receives greater attention in the field of TESOL since testing practices utilized in and associated with language teaching and learning are so much different from such practices in other fields (Brown & Hudson, 1998).

Compared to the traditional ones, alternative assessment procedures are based on ongoing and real life experiences. All methods and techniques of alternative assessment provide various activities which are meaningful and significant to learners, and also include authenticity for learners (Culbertson, 2000). Alternative assessment aims at performing complex data collection, and providing multiple and rich evidence for further judgment (Bachman, 2005). Also, this type of assessment enables the teachers to monitor students' weaknesses and strengths in various contexts (Law & Eckes, 1995). Alternative assessment, which is associated closely with informal, qualitative, classroom – based and performance assessment, is situated in the classroom, helps teachers make their own choices on the measures for the assessment, is based on constructive view of learning and gives importance to the process as well as products (Hoffman, Assaf & Paris, 2001). Alternative assessment is in line with constructivist theory of learning which considers learners as the constructors of their own knowledge. Also, this type of assessment focuses on the entire process of students' learning, allows teachers to monitor students' ongoing progress and helps students engage in self-reflection and self-evaluation through authentic and classroom-based tasks (Janisch, Liu & Akrofi, 2007)

Alternative assessment can be a benefit not only to teachers but also to learners in a variety of ways. Al-Mahrooqi and Denman (2018) assert that alternative assessment allows instruction and assessment to interact continuously which results in a clearer picture of learners' abilities, skills and gains. Furthermore, it consists of performance-based and

authentic tasks which are to be performed in real contexts. Brown and Hudson (1998) list several positive characteristics of alternative assessment in language teaching as; being non-intrusive, including meaningful tasks and instructional activities, being as an extension of everyday life experiences, including authentic tasks, focusing on both processes and products, having transparent criteria and standards, providing information about students' strengths and weaknesses, being sensitive to the learners who have different cultural backgrounds and providing opportunities for teachers to perform new instructional and assessment roles. Similarly, Nasab (2015) also claims that alternative assessment has a number of advantages for teachers, learners and other stakeholders. The benefits of this approach are; to assess the process as well as the product, to scrutinize the instruction, to produce critical outputs for the stakeholders, to help connect other fields, to adopt cooperation within the learning process, to endorse autonomous learning and to support students' psychology. Through the use of alternative assessment, students take an active role in learning processes and participate in real context by being aware of their own strengths and weaknesses (Luoma & Tarnanen, 2003). The collaboration encouraged within alternative assessment practices helps to develop learners' self-efficacy, self-esteem and also intrinsic motivation in learning. (Broadfoot, 2003)

On the other hand, although it has many advantages in teaching and learning processes, alternative assessment may have some challenges and drawbacks when it is misused. Dikli (2003) warns teachers to be aware of these challenges and drawbacks that they may encounter such as subjectivity, reliability, validity of alternative assessment results and also time and energy dedicated to these forms of assessment. Janisch, Liu and Akrofi (2007) divide the obstacles into two categories as external and internal factors. They list the lack of parental and administrative support, giving more emphasis on test scores as external factors; limited English proficiency, inadequate time, resources and money, traditional classroom organization and lack of motivation as internal factors.

Several methods and techniques are presented in the current literature used as an alternative assessment (Brown & Hudson, 1998; Dikli, 2003; Hambleton & Murphy, 1992). Most known methods and techniques used in learning and teaching procedures are as the following; projects, portfolios, conferences, diaries, self- and peer-assessment, journal logs, checklists, videotapes, audiotapes observations.

As given in Figure 1, Bailey (1998) compared the alternative and traditional assessment procedures as below (p.207)

Figure 1

Comparison of Alternative Assessment and Traditional Assessment

Alternative Assessment	Traditional Assessment
<input type="checkbox"/> Continuous, longitudinal assessment	<input type="checkbox"/> On-shot tests
<input type="checkbox"/> Direct tests	<input type="checkbox"/> In-direct tests
<input type="checkbox"/> Authentic tests	<input type="checkbox"/> Inauthentic tests
<input type="checkbox"/> Group projects	<input type="checkbox"/> Individual projects
<input type="checkbox"/> Feedback provided to learners	<input type="checkbox"/> No feedback provided to learners
<input type="checkbox"/> Untimed exams	<input type="checkbox"/> Speeded exams
<input type="checkbox"/> Contextualized test tasks	<input type="checkbox"/> Decontextualized test tasks
<input type="checkbox"/> Criterion -referenced score interpretation	<input type="checkbox"/> Norm-referenced score interpretation
<input type="checkbox"/> Classroom -based tests	<input type="checkbox"/> Standardized tests

2.1.1.5. Criterion-referenced test versus Norm-referenced test. There are two basic ways of interpreting learners' performance: norm-referenced and criterion-referenced tests. Norm-referenced test is to describe the performance in terms of the relative position held in some known group. Criterion-referenced test describes the specific performance that was

demonstrated. Blood and Budd (1972) state that criterion-referenced test is composed of tasks or items for which a clearly definable criterion of successful performance is available. This criterion is stated as a certain proportion or percentage of the items on the test and the norm-referenced test is sometimes called as the test of relative achievement. Its purpose is to place students as accurately as possible somewhere along the range of possible achievement for the test from the very lowest to the highest scores. Mehrens and Lehmann (1973) state, "If we interpreted a score of an individual by comparing his score to those of other individuals (called a norm group) this would be norm referencing. If we interpreted a person's performance by comparing it to some specified behavioral criterion of proficiency, this would be criterion referencing." (p.49)

Wiesnerová (2012) relates the norm referenced test to the society of the classroom and sees this type of interpretation not fair and objective since the criteria depend on the others in the class and any pupil's performance is interpreted based on their peers' performance. On the other hand, Wiesnerová (2012) perceives the criterion – referenced test as to be the most objective since the pupil's performance is interpreted based on the sets of criteria given to them in advance and created together with the pupils.

In the light of the information above, we understand that the criterion-referenced test emphasizes the description of performance whilst a norm-referenced test emphasizes discrimination among individuals. A norm-referenced test is applied if the purpose of the test is to measure the student achievement of instructional objectives so that teachers can rank the students on the basis of the magnitude of these differences. For example, Ayşegül did the exam better than 85 percent of the class. If the purpose of the test is to determine the extend to which the instructional objectives have been achieved and to identify those who have mastered the basic facts, operations and concepts within a subject area and who are ready to move to more advanced subject matter, teachers use and develop criterion-referenced tests.

While constructing norm-referenced test, teachers select the items of average difficulty and eliminate the items that all learners are likely to answer correctly. Because the aim of this is to discriminate, select and group the pupils across relative grouping.

2.1.2. Qualities of instruments used for the assessment. There are certain qualities that every measurement device should possess. The most essential of these are validity, reliability, usability (practicality) and washback.

2.1.2.1. Reliability. There are many definitions for the concept of reliability. Mehrens and Lehmann (1969) define that reliability shows the degree of consistency between two measures of the same thing. Gronlund and Linn (1990) assert that reliability provides a consistency which is necessary for assuring validity and indicates confidence about the results. Anandan (2015) alleges that this term is one of the qualities of tests and describes the reliability as repeatability and consistency of any test. Thus, reliability refers to the consistency of measurement and how consistent test scores or other evaluation results are from one measurement to another.

Many factors may also affect the test and its result and the results' consistency. For example if a single test is administered to the same group in close succession, some variation in scores can be expected because of the temporary fluctuation in memory, attention, effort, emotional strain and guessing. With a longer time period between tests, additional variation in scores may be caused by intervening, learning experiences, changes in health, forgetting...and so on. There are four methods generally used for determining the reliability.

- *Test-Retest Method* - In this method, the teacher administers the same test to the same group on a repeated occasion. A measure of stability is obtained by administering a test to a group of individuals, re-administering the same test to the same group at a later date and correlating the two sets of scores.

- *Equivalent-Form Method* - In this method, the teacher administers two different but equivalent forms of the tests (parallel) to the same group of pupils in close succession. Both of the tests measure the same objectives. After the teacher obtains the results of the two tests, he correlates the results. High correlation indicates that both forms are providing similar results and there is a consistency between two forms.
- *Split- Halve Method (Odd-even method)* – The teacher administers the test to a group of pupils. After obtaining the results, the teacher divides the test into two subtests to score the result. One group consists of odd-numbered items and the other consists of even-numbered items. Every student obtains two scores; the number of the odd-numbered answered correctly and the number of the even-numbered answered correctly. Later the scores obtained from both parts are correlated to obtain reliability.
- *Kuder - Richartson Method (KR20, KR21)* – The teacher uses this method to measure internal consistency. In this method, Kuder-Richartson formula is used to estimate reliability. This method is only applied for multiple-choice items.

2.1.2.2. Factors influencing reliability. Reliability of the assessment and also the instrument used for the assessment have critical importance since it is expected to provide consistency, stability and similarity from one administration to another. However, there are some factors influencing the quality of the test results and assessment procedures in terms of reliability as listed below.

- **Length of Test** - The longer a test is, the higher its reliability will be (Gronlund & Linn, 1990). The longer test will supply more enough samples, and also with the longer test students' guessing factor will be decreased.
- **Spread of Score (group heterogeneity)** - The more heterogeneous the group is, the higher the reliability will be.

- Objectivity - The more subjectively a measure is scored, the lower the reliability will be.
- Difficulty of test - Tests that are so easy that everyone gets all items correct or conversely, so hard that almost everyone gets all the items wrong will have little variability among scores and tend to have lower reliability.

2.1.2.3. Validity. Validity refers to the appropriateness of the interpretations made from test scores and other evaluation results, with regard to a particular use (Gronlund & Linn, 1985). Blood and Budd (1972) explain validity as the consistency with which an instrument measures the variable or variables it is designed to measure. This definition is divided into two parts; the first is consistency of measurement (or reliability), the second concerns the extent to which an instrument measures the variable or variables it is designed to measure. They define the aspect of variables called relevance. If one or the other is absent there can be no validity. The validity of a test, then, is dependent upon its relevance and its reliability. To have some degree of validity, the test must have some degree of both relevance and reliability.

Fulcher and Davidson (2007) state that validity has traditionally been defined as precisely measuring what is to be measured, and that it is appropriate to measure what a test is intended to measure. They further define validity as “...*finding out whether a test actually does measure what is intended.*” (p. 4)

In short, validity is always concerned with the specific use of the results and the soundness of the proposed interpretations.

Validity is a unitary concept based on various kinds of evidence. There are many ways of accumulating evidence to support or challenge the validity of an interpretation of test scores. Fulcher and Davidson (2007) divide validity into three types, each of which demonstrates a test is valid. The ways of accumulating evidence for validity are content-

evidence, criterion-related evidence (including predictive and concurrent validity) and construct evidence. In some sources, face-related evidence is also seen as a type of validity, while it is not by some others (e.g. Fulcher & Davidson, 2007)

- *Content Validity* - Content validity is related to how adequately the content of the test samples the domain about which inferences are to be made. In judging content validity, one should look at both the topics or subject matter covered in the test as well as the type of behavior or task desired from the pupils.
- *Criterion-Related Validity* - In this type of validity, evidences are obtained for a particular test by using a criterion to which it is wished to make prediction (Fulcher & Davidson, 2007). In other words, the main focus is the relationship between a particular test and a criterion which is already validated. Whenever test scores are to be used to predict future performance (*predictive validity*) or to estimate current performance (*concurrent validity*) on some valued measure other than the test itself (called a criterion), we are especially concerned about criterion-related evidence (Gronlund & Linn, 1985).
- *Construct Validity* – This type of validity seeks evidence on the relationship between the theoretical constructs (ideas, notions, concepts) and a specific measuring device (Bachman & Palmer, 1996). Construct validity means whether the instrument or measurement device measures the theoretical construct behind what is to be measured.
- *Face-Validity* - Face validity is related with the appearance of the test. The concept of the problem should be changed by people's interest, that is the problem remains the same, but some phrases change according to people's work and interest. For example, the teacher administers a grammar test to the students. If the teacher is to administer this exam to mechanical engineers, the phrasing in

the test should be related to the concept of mechanical engineering. If the teacher is to administer this exam to a Biology class, the phrasing in the test should be related with biology.

2.1.2.4. Factors influencing validity. The validity of an instrument and also the assessment procedure are influenced by many internal and external factors which lead to the wrong information about the students' progress and learning process. These factors are as follows.

1. Unclear directions
2. Reading vocabulary and sentence structure too difficult.
3. Item difficulty level.
 - Norm-based (average= .50)
 - Criterion-based (level of objective)
4. Poorly constructed item
5. Long and ambiguous structure
6. Time-limit
7. Test too short or too long
8. Improper arrangement of items
9. Identifiable pattern of answers

The factors above affect the tests and test results, and also the validity of the tests. The sources of these factors are the test itself. Some factors also affect tests and test results such as cheating during the exam, unreliable scoring to essay type question, error in scoring and physical and psychological conditions. As seen, many of the factors also affect the test. The sources of these factors are outside the learning environment.

2.1.2.5. Practicality. Practicality, also called as usability, refers to how easy assessment procedures are to administrate. Banchman and Palmers (1996) relate the

practicality to the ease of assessment in terms of design, administration, marking and interpretation of the results. Brown (2004) claims that tests could be considered as to be practical when they are economic and easily administrated, scored and interpreted.

2.1.2.6. Washback. Washback, also called consequential validity (Messick, 1996) or test impact (Bachman & Palmer, 1996), is associated with the impact of testing on learning and teaching processes. (Şentürk, 2013). Individual and / or public examination may have an impact on motivation, tendencies and behaviors of the possible stakeholders; such as learners, parents, teachers and so on. Even though washback is sometimes seen as the impact, it is mainly regarded as the effects of tests on teaching and learning (Şentürk, 2013).

Washback could be examined in terms of its micro and macro influences (Bachman & Palmer, 1996). While micro level influences are more related to the influences of the test within the classroom (e.g. methodologies used by teachers, innovation or change of the curricula and influences of students' learning), macro level influences are more related to the influences of the test within the society (e.g. government, policy makers, school board, parents)

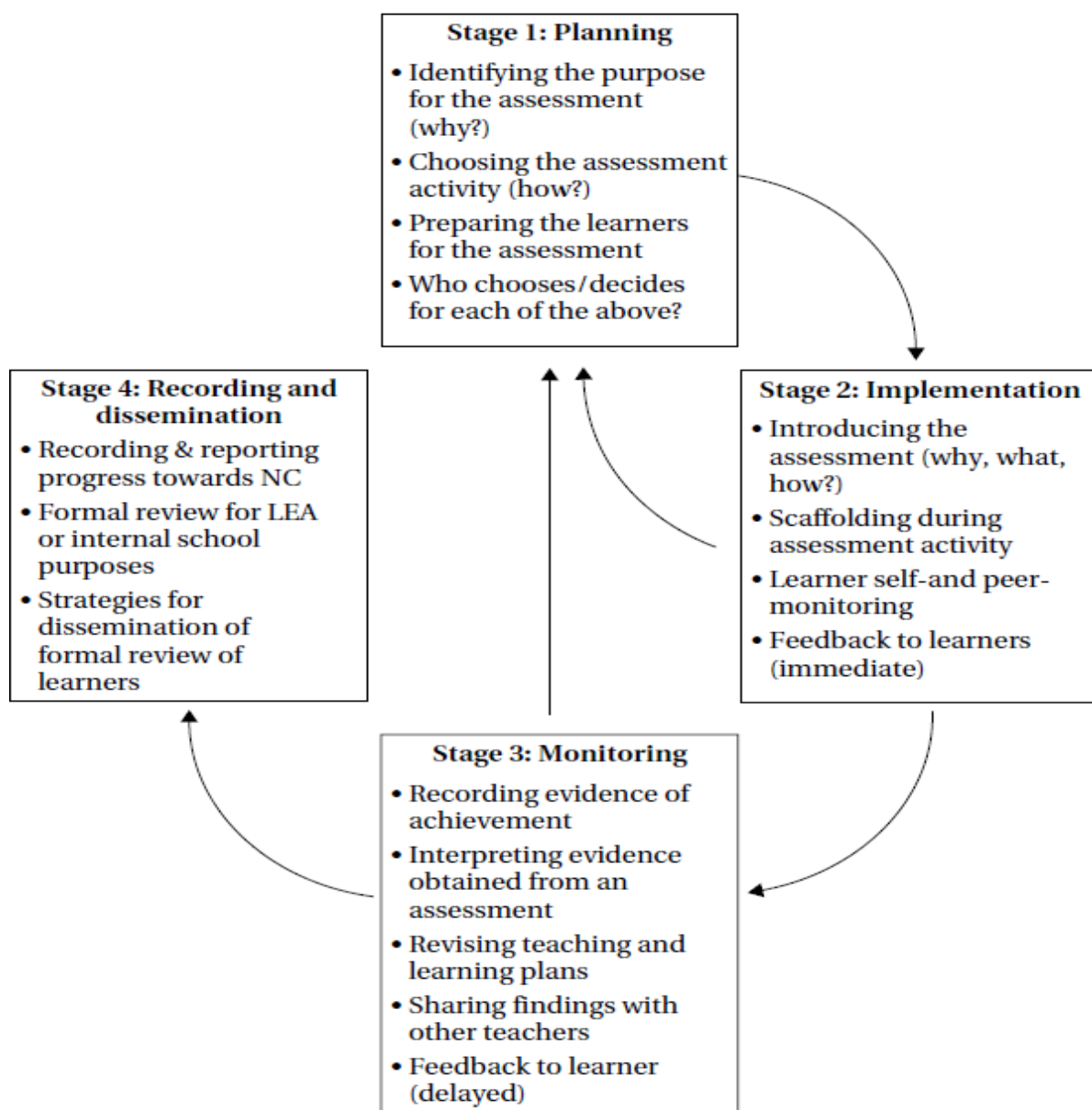
2.2. Assessment in EFL Curriculum and Classes

Assessment in language classes is an inseparable part of language teaching and learning since the information reflected in the assessment is proves the quality of instruction and progress of students in the language. Language classes have a multidimensional context and are social environment in which the students interact with each other and with the teacher, and also involve classroom instructions that directly influence the assessment. In language classes, teachers tend to use classroom-based instructions rather than large-scale assessment since the classes have their own unique characteristics (in terms of students' social and academic background, progress etc.) and assessment in the classroom, which has criterion-

referenced tendency, emphasizes successful learning and successful task completion (Fulcher & Davidson, 2007). Classroom based assessment compared to large scale assessment has a more formative nature (Saito & Inoi, 2017) which results in the continuous assessment of students' progress and provides a lot of evidence through the use of several sources to promote students' learning (Kibar, 2018).

Figure 2

Stages of Classroom-based Assessment (Rea-Dickins, 2001, p. 435)



Classroom based assessment has several stages. Rea-Dickins (2001) proposes four stages of classroom-based assessment based on previous studies, comprehensive literature and observation of teachers. These stages are presented in Figure 1. In the first stage of the classroom-based assessment, called as planning stage, teachers plan the type of assessment, objectives to be attained and also purposes. In the second stage, called as implementation, assessment procedures are introduced to the learners, learners are encouraged to monitor themselves and their peers, and then immediate feedback is provided to the learners. In the third stage, called as monitoring, exam results are interpreted and teachers are involved in self-evaluation for the teaching process. In the fourth stage, called as recording and dissemination, exam results are evaluated and also assessed whether they satisfy the objectives stated in the curriculum.

Assessment in language classes promotes teachers' decision making in their classrooms through providing information not only for teachers to design the instructional process but also for the students to monitor their own progress in developing language skills. Doğandere (2006) remarks that assessment in such classes is undertaken in order to get information or feedback regarding teaching and learning in which teachers could monitor the weaknesses and strengths and also to determine if the learning objectives are attained. This feedback provided for teachers and learners obtained through the assessment procedure makes it significant since it promotes teaching and learning.

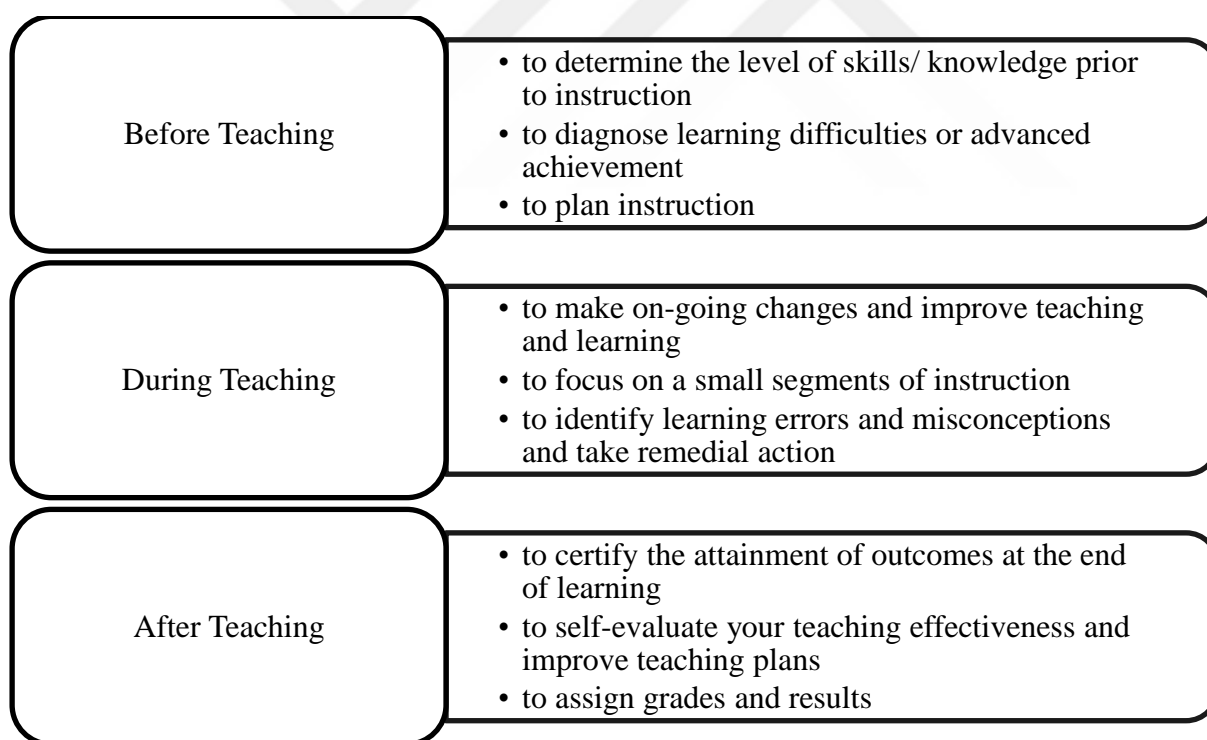
Evidence obtained using educational assessment results help teachers decide on many aspects of educational processes; e.g. decisions on students, decisions on teaching and learning. Teachers' decisions on these processes are critical since they guide students as intended and design the effective implementation based on assessment results. As a result of educational assessment, teachers attempt to respond to some questions and thus make decisions for effective instruction. Teachers intend to respond whether their instructional plan

is appropriate and realistic for the target students, students are ready for the next unit / subject, students are confronted with some difficulties, there are some underachieving and /or advanced students in the class, the instruction is effective and so on. Assessment as a valuable components of teaching process is linked to many educational purposes in English language classes; such as placement, guidance, prediction, diagnosis, selection, evaluation, or other administrative purposes (Lamprianou & Athanasou, 2009).

In the teaching process, classroom assessment is utilized for different purposes at different point / time in the educational processes. Figure 3 summarizes the scope and time of the assessment used for educational purposes.

Figure 3

Scope and Time of Assessment (Taken from Lamprianou & Athanasou, 2009, p. 8)



2.2.1. Teachers' beliefs and practices of assessment. In the educational processes, teachers have several roles to make the process effective and fruitful. One of these roles is associated with the assessment. Cumming (2009) believes that teachers' role for assessment

(as an assessor) is rather complex and consists of many sub-components. In more specific to language teaching, teachers in language classes are generally and routinely expected to monitor and report students' achievement (e.g. written and oral communication skills), diagnose students' learning difficulties, and make student groups for a specific task or placement, etc. Rea-Dickins (2004) says that the teacher role in assessment as a vital part of learning makes the teacher agents of assessment who observes, evaluates and interprets students' achievement for improving their attainment.

In order to monitor the learning process and students' progress and determine whether the pre-determined goals are attained, teachers make use of a variety of assessment procedures and tools. They may prefer to use more teacher-based and standard-based tests, but in this case, the assessment will be more outcome-based and teachers will focus on the final products of the teaching rather than process. In a large scale-testing, the assessor may sometimes not be a teacher of the assessed students and not aware of students' individual differences and also the classroom context, which may result in as a matter of fairness (Fulcher & Davidson, 2007).

On the other hand, in the classroom assessment which is mostly formative in nature and a combination of various assessment procedures, teachers focus more on authentic and informal tasks as well as seek for multi-faced reality. In this type of assessment, teachers (as an assessor) are more familiar with the classroom context, aware of learners' differences and utilized wide range of information / evidences to judge learners' performance. In such assessment, teachers interact with learners for the reason of revealing learners' current ability so as to take a step for further learning (Fulcher & Davidson, 2007). Assessment and teaching are interwoven in alternative assessment, which treats the learning as an active process (Ayas, 2014). In such practices, students are never grouped based on their performance, teachers emphasize what, how and why the students learn, students are given immediate feedback and

teachers look for alternative learning strategies for achieving the learning goals based on the feedback emerged from the assessment (Anderson, 1988).

Figure 4

Assessment for and of Learning (Sutton (2001: 4, as cited in Alberta Education, 2008, p. 4)

Assessment for Learning (formative assessment)	Assessment of Learning (summative assessment)
Checks learning to decide what to do next, then provides suggestions of what to do; teaching and learning are indistinguishable from assessment	Checks what has been learned to date
Is designed to help educators and students improve learning	Is designed for the information of those not directly involved in daily learning and teaching (e.g., school administration, parents, school board, Alberta Education, post-secondary institutions) in addition to educators and students
Is used continually to provide descriptive feedback	Is presented in a periodic report
Usually uses detailed, specific and descriptive feedback, in a formal or informal report	Usually compiles data into a single number, score or mark as a formal report
Is not reported as part of an achievement grade	Is reported as part of an achievement grade
Usually focuses on improvement, compared with the student's previous best	Usually compares the student's learning with other students' learning, e.g., norm-referenced — making learning highly competitive, or the standard for a grade level; e.g., criterion – referenced — making learning more collaborative and individually focused
Must involve the student	Does not always involve the student

Assessment in language classes could be used for different purposes. A discussion on assessment for learning (formative assessment) and assessment of learning (summative assessment) have been observed in the language assessment literature. Sutton (2001: 4, as

cited in Alberta Education, 2008, p. 4) compares assessment *for* and *of* learning as given in Figure 4.

Assessment for learning that associates assessment with formative purposes (so called formative assessment) is mainly undertaken to improve the quality of teaching and learning process, to provide feedback to both students and teaching process and to focus upon improvement and process. Ongoing exchange of information regarding learning between student and student, and also between teacher and students is the main character of this type of assessment (Alberta Education, 2008). On the other hand, assessment of learning that associates assessment with summative purposes (so called summative assessment) is generally used for providing reports on whether students attain the objectives at the end of the course of the subject and focus mainly on grading students' achievements and products. Students' achievement at a given point is assessed or graded in this type of assessment (Alberta Education, 2008).

Similarly, Öz (2014) indicates that the paradigm shift has been observed in the use of assessment in educational practices in which the students' learning and assessment are considered to be interconnected and inseparable from each other. As an integrated part of teaching process, in this new trend, assessment is conceived as a process in order to support students' learning.

Teachers use several assessment tools in language classes to develop learners' language skills and helping them monitor their own learning. At the same time, various usage of assessment tools could help them observe the strengths and weaknesses of teaching and learning process, and also provide evidence for re-designing their instruction. Even though teachers are designing and planning their teaching processes in line with the curricula, a variety of factors influence their choice of assessment types and procedures.

Teachers' beliefs regarding language teaching and learning, their experience in the teaching profession, students' characteristics and also the school types are previously reported as the determinants of teachers' selection or choice of assessment (Broadfoot & Black, 2004; Pajares, 1992). Mertler (1998) reports differences in school levels (elementary, middle and high), years of experience and school location (urban, sub-urban and rural) as the factors impacting teachers' assessment practices. Teachers' perceptions and beliefs in relation to assessment practices have a considerable effect on the ways they used in teaching and their assessment practices as well as teachers' behavior and students' performance (Jia, Eslami & Burlbaw, 2006; Johnson, 1992). In addition, as pointed out in the current national and international literature on language teaching and assessment, teachers' conception and perception of assessment (Gökçe, 2014; Jannati, 2012; Shim, 2009; Öz, 2014) and also attitudes (Brown, 2008) influence their choice of assessment as well as their instructions and assessment practices.

Emphasizing the role of teachers' beliefs in the selection of assessment, Thomas (2012) revealed seven beliefs of teachers on classroom assessment based on the review of available literature. These beliefs are as below;

(1) Assessments which take place informally in the class are the best ways of assessing students' performance, (2) Informal assessments are a waste of teaching time, (3) Assessment is a joint venture between teachers and parents, (4) Assessment encourages students to look critically their own classroom performances, (5) Assessments in the form of direct observation, reduces students' academic achievements, (6) Assessment pressurizes teachers to complete their syllabi, (7) Assessments, in the form of formal tests, makes a negligible contribution to student learning. (p.106)

Song and Koh (2010) claim that teachers' beliefs about student learning often influence their assessment practices. They have conducted a study to examine the relationship

between teachers' beliefs about students' learning and their formative assessment practices and found that teachers' beliefs about students' learning influenced their FA practices; e.g. peer / self -assessment, formative feedback and clarity of task. Their study has also pointed out that teachers' personal interest, tension between assessment of and for learning and their beliefs regarding the feedback and diagnosis of learning needs were the factors behind the teachers' beliefs about students' learning that impacted their choice of FA. Similarly, Guskey (2002) also indicates that teachers' assessment literacy and their beliefs about students' learning influenced mainly their FA practices.

Theory in relation to a specific topic or subject may not be sometimes appropriate and applicable in a particular context due to several reasons, e.g. class size, teachers' experience etc. In this case, teachers tend to select among the available tools based upon their experience and background, course content, students' readiness level and so on. Teachers may have a positive perception of any type of assessment (e.g. formative assessment) but they do not tend to use such assessment due to work load, overcrowded classroom (Büyükkaracı, 2014), difficulty with its administration and lack of time (Veal, 1988).

Assessment in a language classes is a challenging and complicated task and effective use of assessment procedures require knowledge and mastery of assessment strategies. Teachers' perceptions of and beliefs about a particular assessment type are critical for their selection of such assessment, but should not be unique sources. Teachers' knowledge on the assessment and awareness of how the assessment improves students' performance and teaching – learning processes should also be considered (Thomas, 2012).

2.3. Assessing Young Learners in ELT

The term “Young learners” is defined differently by various researchers in terms of the age range. Pinter (2006) defines young learners as the ones between the ages of 3 and 15. McKay (2006) asserts that young learners are the ones studying in primary schools in most

countries and defines young learners as the children between the ages of five and twelve.

However young learners are grouped, they learn, think and study differently from adults. The cognitive development of young learners, in some sources children, is different from adults in terms of their nature of cognition and the timing of the cognitive changes (Britton, 2015).

Each young learner has their own unique characters. Even two children at the same ages have different characteristics. This is because of the fact that young learners have different cognitive, social and emotional characteristics and physical growth at different ages (McKay, 2006; Nunan, 2011) and their social and cultural background as well as their biological maturation are different from each other. Therefore, individual differences of children make them unique. In order to teach English better as a foreign language to young learners and assess their attainment in English, teachers should be aware of their characteristics and tendencies. Being aware of their unique characteristics which are different from adult learners help teachers select the best teaching methods and use appropriate assessment procedures (Doğandere, 2006). Pinter (2006) identifies the characteristics of young learners and differentiates them from the older learners. The differences between young and older learners in terms of their characteristics are summarized in Figure 5 (Pinter, 2006, pp. 2-3).

Figure 5

Characteristics of Young and Older Learners

Young Learners (Aged between 3-15)	Older Learners Aged between (15- older)
<ul style="list-style-type: none"> • Children are at pre-school or in the first couple of years of schooling. • Generally, they have a holistic approach to language, which means that they understand meaningful messages but cannot analyze language yet. • They have lower levels of awareness about themselves as well as about the process of learning. • They have limited reading and writing skills, even in their first language. • Generally, they are more concerned about themselves than others. • They have limited knowledge about the world. • They enjoy fantasy, imagination and movement. 	<ul style="list-style-type: none"> • These children are well established at school and comfortable with school routines. • They show growing interest in analytical approaches, which means that they begin to take an interest in language as an abstract system. • They show growing level of awareness about themselves as language learners and their learning. • They have well-developed skills as readers and writers. • They have a growing awareness of others and their viewpoints. • They have a growing awareness about the world around us. • They begin to show an interest in real-life issues.

Young learners are more concerned about themselves, involved in imagination, fantasy and movement, have a lower level of awareness about learning process and also have a holistic approach to language learning (Pinter, 2006).

Bearing on these differences in mind, all components related to teaching of English to young learners need to be carefully designed and served. Even though teachers put much more effort to develop positive attitude toward learning English and also establish motivating atmosphere for young language learners, they have some difficulties when they start to consider assessment of young language learners (McKay, 2006). Doğandere (2006) states that young learners are full of energy and their concentration could be easily distracted, and thus, if teachers do not use appropriate teaching and assessment procedures in the context of English, young learners' beliefs about learning English may diminish which may result in failure in learning English.

Assessment of young language learners has received greater attention in recent research on teaching English as a second language. Considering their unique characteristics of young learners in English language, a common agreement on assessment tools and procedures to be used for assessing young learners should at least be in line with the following items (Kırkgöz & Ağçam, 2012, p. 121);

- appealing and interesting to the age group, with fun
- using various types of assessment
- using assessment activities to be good learning activities .

McKay (2006) claims that since young learners have their own special characteristics associated with literacy and growth, assessment of young learners should be done carefully by considering two important elements of assessment; validity and fairness. Furthermore, McKay (2006) believes in the effects - positive or negative based on the variety of the factors- of assessment of young learners, and the types of assessment tools and the way the teachers use assessment. In this regard, Clarke (2005) suggests some practical techniques to the teachers to perform assessment for learning in the context of teaching English to young learners. These techniques are; using success criteria, separating the learning objectives from the learning context, focusing on the feedback of the learning objectives and effective questioning. Using wrongly selected assessment tasks may affect some young learners in a negative way.

Assessment of children in language learning should be careful and the evidence attained through the use of assessment should be taken into consideration carefully. Effective use of assessment tools provides feedback to teachers on their teaching and also to young learners about their progress. Parents take a step for moving ahead based on the assessment of their children. Depending on the review of literature on assessment of young learners, Yıldırım and Orsdemir (2013, pp.562-563) report a list of demands to be satisfied for the assessment of young language learners considering their characteristics;

- Tasks should take into consideration children's physical, social, and cognitive development.
- Tasks should be appealing to the age and interest of children and all language should be used in everyday context, matching the way in which they process language.
- Many types of assessment should be used, with the pupils', the parents' and the teacher's perspectives involved.
- Both the tasks and the forms of feedback should underline what they can do instead of what they cannot in order to encourage further learning.
- The pupils should, at least under some circumstances, be given support in carrying out the tasks.
- The activities used in assessment should be good learning practices and serve as a tool to diagnose learning and teaching problems.
- Tasks should be performance-based, requiring students to perform authentic tasks such as giving oral reports, writing essays, cooperative group work, and problem solving.

2.4. Assessment Practices in Elementary Schools

The philosophy behind the curriculum, suggested assessment practices and teachers' perceptions of assessment could diversify the use of assessment tools in EFL classrooms in different countries.

2.4.1. EFL and assessment practices in elementary schools in Turkey. Teaching English as a foreign language starts at 2nd grade as compulsory in primary education in Turkey. English language curriculum for the level of primary schools (2nd to 8th grades) has undergone the reform attempts in terms of theoretical framework in general and in grade level in 2018 (MEB, 2018). The rationales behind the curriculum renewal are associated with a need for providing high-quality English education by up-dating an effective curriculum as

well as for accommodating the curriculum to the educational model (4+4+4 system). The principles of Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) are considered while re-designing the English language curriculum for primary schools. Of the language skills, only listening and speaking skills are focused at 2nd grade in A1 level. In the 3rd and 4th grade, in addition to listening and speaking skills, a very limited emphasis is given to reading and writing in A1 level. The main activities suggested for these grades are TPR, arts and crafts, and drama. For the following grades, the focus on four language skills is shown in Figure 6.

Figure 6

Skills and Suggested Activities by Grades in 2018 English Language Curriculum in Turkey (MEB, 2018, p.10)

Levels (CEFR) (Hours / Week)	Grades	Skills Focus	Main activities / Strategies
	2	Listening and speaking	
1 [A1] (2)	3	Listening and speaking Very limited reading and writing	TPR / Arts and crafts / Drama
	4	Listening and speaking Very limited reading and writing	
2 [A1] (3)	5	Listening and speaking Limited reading Very limited writing	Drama / role play
	6	Listening and speaking Limited reading Limited writing	
3 [A2] (4)	7	Primary : Listening and Speaking Secondary : Reading and Writing	Theme - based
	8	Primary : Listening and Speaking Secondary : Reading and Writing	

Assessment is one of the most important aspects of EFL primary school curriculum in Turkey. The curriculum suggests English teachers to use various types of assessment techniques. In the curriculum, alternative types of assessment techniques are primarily suggested along with self-assessment techniques. In addition to process-oriented testing and self-assessment which encourage students to monitor their own progress in English, formal evaluation including quizzes, written and oral exams, projects and homework assignments are also suggested to be carried out for the assessment of students' success (MEB, 2018).

Assessment of 2nd and 3rd graders is not carried out for summative purposes but formative purposes. On the other hand, summative and formative assessments together are offered for 4th graders and onwards in order to test students' language proficiency. Suggested alternative assessment techniques are portfolio assessment, project assessment, performance assessment, creative drama tasks, class newspaper / social media project, journal performance and etc.

2.4.2. EFL and assessment practices in elementary schools in Italy and Finland. In Europe, the Common European Framework of Reference for Languages: Learning, Teaching and Assessment (CEFR) has been considered for level of language proficiency and assessment procedures so as to design language curriculum. CEFR provides a common guideline for language teaching in terms of curriculum guidelines, syllabus, textbooks and examination. In CEFR, an action-oriented approach is adopted in order to develop language competencies (sum of knowledge, skills and characteristics) in general competencies and in specific communicative language competencies (Council of Europe, 2001). CEFR suggests levels of language learning on a global scale of education. The levels of languages learned are A1 and A2 for basic users, B1 and B2 for independent users, and C1 and C2 for proficient language users. The common reference levels suggested by CEFR are given in Figure 7.

Figure 7

Common Reference Levels (Council of Europe, 2011, p. 24)

		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.
	C2	Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
PROFICIENT USER		Can understand a wide range of demanding, longer texts, and recognise 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 organisational patterns, connectors and cohesive devices.
	C1	
	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. 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.
INDEPENDENT USER		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 & ambitions and briefly give reasons and explanations for opinions and plans.
	B1	
BASIC USER		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.
	A2	

A1	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 CEFR, rather than specific assessment tools and procedures, different assessment procedures are suggested by considering validity, reliability and accuracy of decision for assessing proficiency of language users. The answers to two critical questions “*what is assessed?*” and “*How is performance interpreted?*” are the central to assessment procedures suggested in CEFR. For particular aspects of proficiency, two general types of assessment are suggested as self- or teacher assessment, and performance assessment. The types of assessment suggested for assessing learners’ proficiency are demonstrated in Figure 8.

Figure 8

Types of Assessment Suggested in CEFR (Council of Europe, 2011, p.182)

1	Achievement assessment	Proficiency assessment
2	Non-referencing (NR)	Criterion – referencing (CR)
3	Mastery learning CR	Continuum CR
4	Continuous assessment	Fixed assessment points
5	Formative assessment	Summative assessment
6	Direct assessment	Indirect assessment
7	Performance assessment	Knowledge assessment
8	Subjective assessment	Objective assessment
9	Checklist rating	Performance rating
10	Impression	Guided judgment
11	Holistic assessment	Analytic assessment
12	Series assessment	Category assessment
13	Assessment by others	Self- assessment

European countries primarily follow these guidelines in designing and implementing language teaching programs for various levels of education even though teachers make use of suggested teaching activities and assessment tools based on their context of teaching and many other factors (e.g. classroom, students, etc.). Thus, CEFR provides common framework and suggests common activities, but teachers have high level of autonomy and contextualize them in their classroom conditions depending on their philosophy of the national curriculum.

Learning two foreign languages is compulsory for all students starting at the primary school levels in Italy. Among other languages, English as a second language is mandatory at the age of 6 in the primary schools. Compared to other European countries, Italy is one of the countries starting at English language teaching earliest in education system (Faez, 2011). English lesson hours in a week are varied in different classes in primary education such as one hour for year - 1, two hours for year - 2, and three hours for year – 3 to 5 (Hayes, 2014). Students' competencies gained in English are continuously assessed by the use of teacher-assessment and self-assessment procedures according to European Framework levels. The assessment results are generally turned into credits which are included in students' portfolios (Lopriore, 2002).

Language teaching starts at 3rd grade in primary schools in Finland. In the language education, predominantly English teaching, the objectives for grades 3 to 6 are more associated with having communication skills, first oral skills and then writing skills (Hayes, 2014). A1 syllabus for English or other foreign language is applied to 3rd grade, A2 syllabus for English or other foreign language is applied to 4th and 5th grade. B1 syllabus for 7th grade, B2 for 8th and 9th grade, and B3 syllabus for upper secondary schools are applied in Finish schools. English is one of the most chosen languages among the students (Helminen, 2013).

2.5. Research on Young Learners in Foreign Language Classes

Teachers' use various assessment procedures, their perceptions and their assessment practices in teaching English as a foreign language for young learners are observed in many research studies in different countries. However, the current literature reveals a limited number of assessment studies done in the context of Italy and Finland when compared to Turkey. This is due to the fact that the issue of assessment may not be the main focus of research studies in Italy and Finland or the language of published studies on the assessment were written in Italian and Finnish so that they are not available in the literature written in English language.

In this part, firstly, research on teachers' assessment practices and preferences in EFL classes in three countries (Turkey, Italy and Finland) and then in other countries are presented. Later, research on teachers' beliefs about assessment and assessment practices are summarized.

2.5.1. Assessment practices in foreign language classroom in Turkey. Kırkgöz and Ağçam (2012) examined Turkish EFL teachers' written assessment practices in the 4th and 5th grades in primary schools by comparing the question types before and following the 2005 English Language Teaching curriculum. For data sources, a total of 100 written exam papers taken from 25 EFL teachers working at public primary schools were content analyzed across previous categories suggested in the literature. The analysis of the written exam papers revealed that test items such as sentence completion, transformation, translation and responding to questions were used more frequently in the exams in both grades after 2005 ELT curriculum.

Ayas (2014) conducted a study to investigate English teachers' conceptions regarding the assessment and choices of assessment tasks. As the data collection instrument, Teachers' Conceptions of Assessment Questionnaire (TCoA-III A) was used. The instrument was

applied to 43 teachers teaching English as a foreign language in primary schools. The results of the study pointed out that English teachers tended to use various assessment tools and strategies in practice. Most of them preferred to use alternative assessment procedures even though some preferred to use traditional ones. No contradiction was observed between teachers' perceived preferences and conceptions regarding assessment. Interestingly, even though there seemed to be no contradiction, the teachers' actual use of assessment procedures was observed to be different from their perceived preferences and their conception of the assessment. The assessment tools that teachers used in practice were teacher-made written test, standardized tests, question – answer, written works by students, self- and peer-assessment, portfolios, and planned observation. Teachers conceived the assessment in three categories as improvement, student accountability and school accountability.

Öz (2014) examined Turkish EFL teachers' assessment for learning practices and preferences of assessment methods in EFL classes. He also examined whether teachers' practices of assessment or learning differed with regard to gender, years of teaching experience, school type (public vs. private). 120 Turkish EFL teachers participated in the study and filled out an online questionnaire. The results pointed out that EFL teachers most frequently preferred to use fill-in-the-blank, multiple choices, true-false, matching and short-answers as assessment methods. They more tended to implement traditional methods rather than alternative ones (e.g. portfolios, performance assessment and so on.). Teachers' scaffolding practices and perceived monitoring were reported to be as high level. Teachers' assessment practices in relation to monitoring in support of student learning significantly differed with regard to gender, school type (public vs. private) and years of teaching experience.

Kirkgoz, Babanoğlu and Ağçam (2017) investigated EFL teachers' perceptions of assessment and also their use of assessment and question types in assessment practices in their

language classes in primary schools. A total of 56 EFL teachers drawn from 42 public primary schools were selected as the sample of the study. In order to elicit EFL teachers' perception and practices related to foreign language assessment in 4th to 6th grades, a questionnaire was used as a data collection tool. The results revealed that nearly all of the teachers believed in the importance of using assessment in foreign language classes for the reason of tracking the development of students in foreign language, examining the understanding of students in subject matter in foreign language, determining whether the objectives in the course curricula were attained, evaluating the performance of both the teacher and students, increasing the quality of teaching, giving feedback not only to the teachers for their teaching but also to the students for their progress and also assessing the appropriateness of approaches, methods and techniques used in language teaching and suggesting necessary changes (if needed) for the implementation. The most frequently assessment tools used in language classes were paper-pencil tests, in-class observations and performance tasks. In order to evaluate the students' attainments better in English language, teachers preferred to use performance-based and communication-based assessment types more frequently than traditional assessment types. Oral exams, presentations, portfolios, projects and quizzes were less frequently used assessment tools in English language classes.

Kibar (2018) conducted a study aiming at determining the differences between perceptions of in-service and pre-service teachers on classroom-based language assessment. A total of 40 in-service English teachers and 99 pre-service English teachers in Ankara participated in the study. The participants were asked to complete a questionnaire to reflect their perceptions on assessment in language classes. The results pointed out that there was a significant difference between two groups of teachers regarding classroom – based assessment. The participants were observed to have positive perceptions with regard to classroom-based assessment. Pre-service teachers' teaching experience made a difference on

the teachers' perceptions. On the other hand, the pre-service teachers perceptions of classroom based assessment did not significantly differ in terms of their gender and grade point average (GPA).

Önalın and Karagöl (2018) examined teachers' beliefs about the use of assessment in EFL classes, and also the effect of teachers' years of teaching experiences and the departments they graduated. 70 Turkish EFL teachers selected from prep school of a university completed a questionnaire. The results of the study pointed out that Turkish EFL teachers mostly tended to use assessment for formative purposes. Self-assessment was believed to be an important assessment procedure in teaching English as a foreign language. Teachers' beliefs about the use of assessment did not significantly differ in terms of neither years of teaching experience (novice vs. two or more years) nor the department (ELT vs. others) graduated.

2.5.2. Research on assessment practices in foreign language classroom in Italy.

Gattullo (2000) undertook a study with four teachers, three generalists, one specialist and also 86 children (at the age of 9 and 10) studying English as a foreign language since they were 8 years-old. In this study, Gattullo (2000) aimed at identifying interpretation and implementation of assessment in the final years of primary schools, dimensions of formative assessment, good practices used as formative assessment and differences between formative and summative assessment. For data sources, classroom observations, interviews with teachers, assessment materials used for teaching English as a foreign language and student questionnaire were utilized. The results pointed out that the teachers used formative assessment for the purpose of correcting, questioning, metacognitive questioning, judging, observing process, examining products, rewarding, clarifying, task criteria. Questioning in formative assessment was used for two purposes; correcting or making counter-suggesting and providing input, response and feedback. Several weaknesses in using formative assessment in classroom teaching were

revealed in the observations, interviews and questionnaires as the development of more structural ways in oral abilities, providing action and feedback for self-repair and developing procedures in order to record students' performances.

2.5.3. Research on assessment practices in foreign language classroom in Finland.

Wewer (2014) designed a mixed design study to investigate teachers' assessment practices in second language classes in primary schools in Finland. In addition, parents' and pupils' perception of language assessment and computer simulation program (LangPerform) were also examined in the study. A total of 42 teachers serving bilingual content instruction, 109 children drawn from 3rd to 5th grades and also 99 parents participated in the study. Mixed type data were collected; experimentation, questionnaire with children and parents, and video interview were used as data sources. The results of the study showed that the most commonly used assessment tools in language classes were teacher observation, dialogic interaction and bilingual tests. On the other hand, the least used assessment tools were peer assessment, simulations and portfolios. Many of them used assessment in order to gather data on the students' attainment and only one third of the teachers scarcely provided feedback to pupils based on the assessment even though teachers perceived the language assessment important.

2.5.4 Research on assessment practices in foreign language classroom in other countries.

Chan (2008) examined teachers' practices and beliefs about multiple assessment and their difficulties while practicing the assessment in English language teaching in Taiwan. A total of 520 EFL teachers selected from elementary schools in Taiwan participated in the study and completed a questionnaire including various types of items. The results of the study showed that teachers' beliefs and practices of assessment differed in terms of their years of ESL teaching experiences, but did not differ in terms of their ages. Teachers believed that understanding students' learning success and their progress were two main purposes of the assessment. The most frequently used assessment types used by the teachers were paper-

pencil tests, classroom observation and task-based assessment. None of the teachers selected traditional types of assessment as the main assessment types. Many of the teachers (%70) used alternative and traditional types of assessment together. The obstacles influencing teachers' practices of assessment mostly were heavy curriculum content, classroom size (crowded classroom) and considering the assessment as time consuming.

Shim (2009) investigated Korean teachers' perceptions and the use of classroom-based assessment in the class where English was taught as a second language in Korean primary schools. In the study, the author examined teachers' assessment in four stages as planning, implementation, monitoring and recording – dissemination. For data collection, a questionnaire including both qualitative and quantitative items reflecting the stages of classroom-assessment was utilized. The results revealed that teachers' firm beliefs about the classroom assessment were influential in their assessment practices, their knowledge of assessment was in a good level and they used appropriate assessment procedures in their operation of classroom assessment practices in the context of English teaching. Although English teachers held many principles in terms of classroom-based assessment, they did not put all into the practice due to heavy work-load, limited funding for teaching foreign language, central bureaucracy of the educational system and overcrowded classroom size. Furthermore, teachers' assessment practices were also affected by their complex relationship with head teachers, other teachers and also parents.

Jannati (2012) examined Iranian ELT teachers' perceptions and practices in relation to assessment in language classes. A total of 18 ELT teachers in English language institutes in Iran participated in the study. Teachers in the study were divided as low, mid and high in terms of their teaching experience levels. Interviews with teachers were used as data sources in order to address to the aim of the research. Content analysis of the interview transcript pointed out that all teachers were observed to be aware of basic concepts and technologies

associated with the assessment in language classes. Teachers' perceptions of assessment were not differentiated with regard to their teaching experience. Teachers at different level of teaching experience attributed the characteristics of good assessment in a different way. The most frequently and commonly reported features were that assessment should assess students' achievement (whether they learn what is expected to be learned) and assessment should be designed based on the objectives of the courses. The assessment practices the teachers used in English language classes were asking questions, asking students to write, storytelling, role-play, group-work and pair-work.

Britton (2015) designed a mixed-method research study to investigate EFL teachers' understanding of assessment for learning in primary schools, implementation of their assessment and the impact of their implementation on the interaction in the class. Various data sources such as class observation, individual and focus group interview and also school documents were utilized. Content analysis of the data showed that teachers' implementation of assessment in the classes of 7-9 year old students and 10-11 year old students were different in terms of setting objectives and expectations, monitoring performance and checking achievement. In addition, empirical data revealed that teachers' use of assessment for learning could have an impact on creating a classroom environment conducive to learning on collaborative interactions. Teachers believed in the special characteristics of teaching English to young learners and their understanding of assessment for learning was in line with theoretical frameworks taken place in the literature.

2.5.5. Research on teachers' perceptions of and beliefs about assessment. Jia, Eslami and Burlbaw (2006) designed a qualitative study to examine English teachers' perceptions of classroom-based assessment in reading. In their study, they also aim at presenting teachers' use of classroom based assessment in reading, their understanding in relation to the effectiveness and function of this type of assessment and also the factors

associated with this type of assessment. 13 teachers teaching English as a second language in middle (six teachers) and elementary (seven teachers) schools participated in the study. For data sources, classroom observations, interview with English teachers and teachers' assessment materials were used. Content analysis of the data pointed out that teachers valued classroom-based assessment in reading, they believed this type of assessment was valuable and accurate, and also it had high contribution to reading instructions in English classes. As compared to classroom-based assessment, the teachers perceived standardized testing in a negative way and placed very little value in English classes. The factors influencing the use of teachers' assessment in reading were district policies, state-mandated tests and also students' characteristics.

Şahinkarakaş (2012) examined EFL teachers' conceptions of assessment and the effect of teaching experience on their conceptions of assessment. A total of 53 prospective teachers in the department of English language teaching and 47 EFL teachers having a lot of teaching experiences participated in the study. Teachers were given a statement to indicate their metaphorical images for assessment and then required to give their reasons for this metaphor. The results of the study showed that teachers' produced 73 metaphors categorized in four themes; assessment as a formative tool, a summative tool, something agitating and a sign of self-efficacy. The metaphors were categorized under the assessment as a formative tool were related to an ongoing process, a way to promote learning through feedback and diagnosing learners' weaknesses and strengths; the metaphors categorized under the assessment as a summative tool were related to the end product and grading students' products; the metaphors categorized under assessment as something agitating were related to fear and difficulty; and the metaphors categorized under the assessment as a sign of self-efficacy were related to reflection of oneself and personal satisfaction.

Gökçe (2014) examined EFL teachers' perceptions of formative assessment (FA) and also compared their perceptions of FA in terms of school type (public vs. private). A total of 100 EFL teachers were selected from 39 schools (primary to high school) in Turkey. The results pointed out that most of the EFL teachers emphasized the importance of using FA due to its contribution to students' learning in English. Furthermore, the teachers believed in that the students should be aware of why they are being assessed and FA helped promote learning in language classes. More than half of the EFL teachers in public schools emphasized the value of FA that promotes learning, they benefited from FA in order to re-design and develop their own teaching of English. Private school EFL teachers also believed in that the primary aim of FA was to promote learning. They claimed that students could progress on their own. The data also revealed that the private school EFL teachers used FA more effectively than those in public schools did.

Halali, Singh, Saeed and Making (2017) designed a survey study to find out EFL teachers' perceptions of using classroom-based assessment in the context of second language acquisition of reading class. Drawn from elementary schools, 28 ELT teachers having ESL certified and at least one year teaching experience participated in the study. The data were collected from the self-reported questionnaire developed by the researchers. The results of the study showed that there were several internal and external factors greatly influencing teachers' perceptions of using classroom-based assessment. Of the factors, facilities, willingness, time management, suitable textbooks and the support provided by the ministry were more observable ones. Teachers believed that using classroom-based assessment helped to improve their formal instruction. Positive internal and external factors promoted teachers' motivation toward classroom-based assessment in reading.

Chapter 3

Method

In this part, research design, participants and their characteristics, data collection instruments, data collection and analysis procedures are presented in detail.

3.1. Research Design

Mixed methods research was utilized to find out EFL teachers' beliefs and practices with regard to the assessment of young learners (4th graders) and to discover whether there are matches or mismatches between the expected quality of assessment and the reality in Turkey, Italy and Finland. The research focused on 4th grade level because formal assessment starts at 4th grades in Turkey. Various aspects of the assessment system (types, activities, content, skills etc) were analyzed in detail and comparatively in 3 countries to find out the problems and deficiencies. Using qualitative and quantitative procedures together enabled to understand teachers' assessment practices in 4th grade EFL classes in the selected countries. Mixed methods research integrates qualitative and quantitative research paradigms and engages in collecting, analyzing and interpreting qualitative and quantitative data together to provide a general picture addressing to the research problem (Creswell, 2003). Mixed methods research includes further sub-models. In the present study, first quantitative data and then qualitative data were collected which enabled to explain general tendencies of EFL teachers. It was named as explanatory sequential design (Creswell & Plano Clark, 2011) referring to collection and analysis of data in quantitative nature followed by qualitative data (Terrell, 2011). This type was preferred because the study aimed to explore and explain the assessment practices at 4th grades with EFL teachers' self-reports (questionnaires) and in-depth views (interviews).

Mixed methods research enabled to undertake the study in two stages successively. In the first stage, collecting data through the use of a questionnaire in a quantitative paradigm

enabled the researcher to collect data about the current use and practice of assessment in classes of the selected countries. A questionnaire, beyond a list of questions / items, requires very serious work to design, develop and administer. A questionnaire is a self – reported research instrument utilized to collect information from huge amount of people so as to describe peoples' life conditions, behaviors, beliefs or attitudes in a short period of time (Akbayrak, 2000; Thomas, 1998). As compared with other data collection tools (e.g. observation, interview etc.), a questionnaire has advantages such as collecting data very quickly from larger groups from different regions with lower costs. On the other hand, inflexibility of the items (pre-designed items / questions), social desirability effect, difficulty in collecting data on some sensitive issues (e.g. religion, birth control, political tendencies etc.) are perceived as the disadvantages of the questionnaires (Büyüköztürk, 2005). Akbayrak (2000) asserts that the main advantages of a questionnaire are related to anonymity, application skills, biases, confidentiality, cost, data analysis, time, sample size and sampling whereas the disadvantages of the questionnaire are related with access to information, flexibility, response rate, reliability and validity.

In the second stage, collecting data through the use of interview with teachers in qualitative paradigms enabled the researcher to present the teachers' view on assessment procedures (practices, difficulties, suggestions etc.) they used in their classes. At the same time, this stage helped the researcher to understand the feelings, beliefs, views behind their tendencies reported in the first stage. The interviews, as a data collection tool, are a kind of conversation between (at least) two people for a specific purpose to collect data for (a) specific research purpose(s) (Akbayrak, 2000) and also considered to be “...*interchange of views between two persons conversing about a theme of mutual interest*” (Kvale, 1996. 14). There are mainly three types of interviews as structured interview, semi-structured interview and open ended interview conducted as face-to-face individually or as a focus group. In the

study a semi-structured interview was used since it enables the researchers to ask pre-determined questions with further prompt questions for deepening the conversation. The interviews have several advantages as well as some limitations. Obtaining in-depth information in order to discover things (case events, phonemes etc.) and probing interviewee's responses to gain deeper understanding and more detailed information are the two main advantages of the interviews (Hobson and Townsend, 2010). In addition, Akbayrak (2000) asserted that advantages of the interviews are related to ease of accessing to information, flexibility, reliability, validity and response rate. On the other hand, she believed that anonymity, application skills, biases, confidentiality, cost, data analysis, time, sample size and sampling are the main disadvantages of the interviews.

3.2. Participants

A total of 89 English teachers in Turkey, 40 English teachers in Italy and 41 English teachers in Finland constituted the sample of the study. Teachers who participated in the study were the ones teaching at 4th grade English class at primary schools and having former experiences at primary and pre-primary schools. Due to the researcher's accessibility to EFL teachers in Italy and Finland through Erasmus+ projects, they were included in this study. Teachers' participation in the study was based on voluntariness and teachers' ages, experiences, types of schools and such variables were not considered on including them into the study. Therefore, their convenience was the main concern of the inclusion. Teachers were accessed through the use of social networks and teachers' forums. Demographics of English teachers are given in Table 1.

Table 1

Demographic Characteristics of the Participants (N = 170)

	<u>Turkey</u>		<u>Italy</u>		<u>Finland</u>	
	f(n)	p (%)	f(n)	p(%)	f (n)	p (%)
<u>Gender</u>						
Male	16	17,79	2	5	-	0
Female	73	82	38	95	41	100
<u>Teaching Experience</u>						
0 to 5 years	11	12,35	9	22,5	11	26,82
6 to 15 years	61	68,53	16	40	14	34,14
16 to 24 years	15	16,85	12	30	13	31,70
24 years or more	2	2,24	3	7,5	3	7,3
<u>Level of Education</u>						
Bachelor's Degree	79	88,76	23	57,5	7	17
Master's Degree	10	11,23	17	42,5	34	82,9
<u>School type</u>						
Public	86	96,62	30	75	36	87,8
Private	3	3,37	10	25	5	12,19

Of English teachers selected from Turkey, 16 were male, 73 were female; 11 teachers had 0 to 5 years teaching experiences, 61 had 6 to 15 years teaching experiences, 15 had 16 to 24 years teaching experiences and 2 had 24 years or more teaching experiences; 79 teachers had bachelor's degree and 10 had master's degree; and 68 worked at public schools and 3 worked at private schools. Of English teachers selected from Italy, 2 were male and 38 were female; 9 teachers had 0 to 5 years teaching experiences, 16 had 6 to 15 years teaching experiences, 12 had 16 to 24 years teaching experiences and 3 had 24 years or more teaching experiences; 23 teachers had bachelor's degree and 17 had master's degree; and 30 worked at public schools and 10 worked at private schools. Of English teachers selected from Finland, all were female; 11 teachers had 0 to 5 years teaching experiences, 14 had 6 to 15 years teaching experiences, 13 had 16 to 24 years teaching experiences and 3 had 24 years or more

teaching experiences; 7 teachers had bachelor's degree and 34 had master's degree; and 36 worked at public schools and 5 worked at private schools.

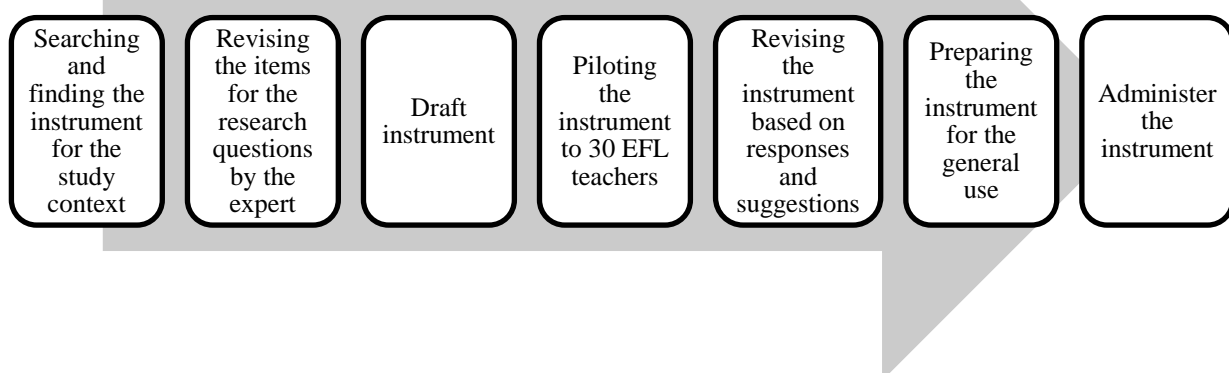
3.3. Data Collection Instruments

Two instruments were used for the data collection. The first one was a questionnaire and the second one was a semi-structured interview.

3.3.1. Questionnaire. A questionnaire, developed by Brumen et al. (2009), was revised and adapted into the study context to identify teachers' beliefs and practices on the assessment of 4th grade EFL students. In the revision of the original instrument for adapting it to the study context, several steps were carefully taken. The Figure 9 summarizes the steps followed.

Figure 9

The Steps Followed for Revision and Adaptation of the Instrument



First of all, the original instrument was drawn from the internet based on its most appropriateness to the purpose of the study and examined item-by-item to determine to what extent the items in the instrument best fit the study's aim and whether a revision was needed over the instrument. Then, the first revision of the instrument was done by the expert on the assessment of young EFL learners. The revision was undertaken in terms of language

(understanding), format and appropriateness of the items to the study aim. Thus, some items were found to be irrelevant to the study and excluded from the original form, some items were revised by adapting them into the study context and some items' formats were changed (e.g. from "tick" to "yes – no" response). Furthermore, this latest format (draft) of the instrument was applied as a pilot testing to 30 Turkish EFL teachers, teaching 4th graders at primary school. The purpose of the pilot testing was to give the instrument the final version. The teachers in the pilot test were asked to examine the items with regard to their fitness to the purpose of the study and provide any revision / change over the items if needed. Time items in the instrument were also revised and changed based on the responses of the teachers and their suggestions on the items. The pilot testing indicated the necessity to revise some items in terms of format, content and structure. Some of the examples related to the revision from the first draft to general use are given below.

Revision based on expert opinion (draft)

7. Why is assessment carried out in your 4th grade-classroom?

- To provide information for national FL standards (school curriculum, goals)
- To provide information to parents
- To provide information to pupils
- To provide information to you as a language teacher
- As a part of certification at the end of primary school
- To provide information for outside evaluators

Other, please specify: _____

8. Who is responsible for the preparation and performance of assessment of 4th grade students?

- A language teacher, who teaches young learners
- A group of FL teachers within the school
- Outside the school; national education authority
- Local education authority

Other, please specify: _____

Revision based on pilot testing (general use)

7. Why is assessment carried out in your classroom?

- To check students strength and weaknesses for lesson planning
- To provide information to parents and students
- To place students
- To motivate students to learn
- To provide a record of students' achievement for success and failure

Other, please specify: _____

8. Who is responsible for the preparation of assessment?

- A foreign language teacher
- A primary teacher
- A group of language teachers within the school
- Local education authority
- National education authority

The final version of the instrument was designated based on the revision provided by the teachers in the pilot testing. The final version included two main parts with different types of items. The first part was designed to determine the characteristics of teachers' assessments in practice. The first five items are regarded as demographics of the teachers. Next items were designed to assess teachers' beliefs about assessment, the purpose of the assessment which the teachers use for 4th graders, the responsible people for the preparation of assessment, teachers' beliefs and practices in terms of reporting results, teachers' trainings on assessment, the types of assessment used in 4th grade EFL classes, the content / skills – focus in the assessment of students and the kinds of exercises used in the classes. The second part was designed to find out teachers' beliefs about assessment. This part has four sub-sections as beliefs about the types of assessment that teachers use for 4th grade students, teachers' content / skills – focus in the assessment of 4th grade students, kinds of exercises teachers used in their assessment of 4th grade students and assessment of students' foreign language. Therefore, the questionnaire consists of 5 demographics (gender, country, teaching experience, level of education and class size) and 15 questions / sub-sections requiring a variety of responses (e.g. Likert type, yes – no, select more than one, open-ended etc.). The questionnaire is given in Appendix 1.

3.3.2. Semi-structured interview. Semi-structured interviews were used to investigate English teachers' beliefs about and practices on the assessment procedures used in 4th grade English classes, their difficulties and problems during assessing young learners' attainments in English, and also suggestions for improving assessment system in Turkey, Italy and Finland. The interview schedule included 10 open-ended questions addressing to teachers' beliefs about exams and a lot of assessment procedures for young learners, teachers' use of and practices on various assessment procedures assessing young learners' skills, teachers' difficulties in assessing young learners, and teachers' suggestions for improving

their own assessment skills and also for assessing young learners. The interview is given in Appendix 2.

3.4. Data Collection Procedures

Both instruments were administrated successively in the fall semester of 2018 - 2019 academic year. Firstly, the necessary permission to undertake this study with EFL teachers at primary schools was obtained from Provincial Directorate of National Education in Bursa, Turkey. This permission was only taken for Turkish EFL teachers. For Italian and Finnish EFL teachers, no official permission was obtained. These teachers were contacted through internet and participated in the study based on their voluntariness. After obtaining permission, the questionnaire was administered to 30 English teachers in Bursa as a pilot testing. Having the questionnaire revised, it was administrated to the selected teachers face to face in Turkey, and via teacher networks and social media in Italy and Finland. Finally, the interview was undertaken based on teachers' availability with 9 teachers in Turkey, 8 teachers in Italy and 6 teachers in Finland. The ones took part in the interview were drawn from those who filled out the questionnaire. The interview schedule was sent to the teachers via internet and asked the teachers to respond open-ended items in the schedule.

3.5. Data Analysis

Since both the qualitative and quantitative data were collected in the study, separate analysis procedures were followed. For the quantitative part of the study, the data collected through questionnaire were firstly entered into SPSS 23. The data were cleaned in terms of missing cases, outlier etc. Later, the cleaned data were analyzed by making use of descriptive statistics. The results obtained through quantitative analyses were presented using frequencies, percentages, means and standard deviation along with the tables which enabled to make comparison among the countries. In addition to descriptive statistics, one-way-

ANOVA was performed for the comparison of data obtained from teachers from three countries.

For the qualitative part of the study, the interviews (written documents) were subjected to content analysis in order to reach common codes and themes. The analysis of the data was done manually. The qualitative results were presented using emerged codes and themes. The direct quotation from the teacher interviews were also used to present the codes and themes better. Teacher's anonymity was taken into account and direct quotation taken from teacher interview was given together with the codes rather than real names. For example, their names were coded as T_EA, I_CF and F_TW. The first letter refers to the country (T-Turkey, I-Italy and F-Finland), the other letters refer to the initials of teachers' real names. In order to assure the content analysis reliability, several steps were followed. Firstly, all interview transcripts were coded by a researcher who was an expert on assessment of young learners and qualitative coding. Later, a code list emerged from this initial coding of all interviews was given to another researcher who was an expert on the same topics. This independent researcher was asked to examine the codes across the interview transcripts and later indicate whether the codes reported in the initial coding were really inherent to the interview transcripts. Nearly all of the codes emerged in the initial code list were almost confirmed by the second researcher which meant that the codes were consistent from the first coding to the second one.

Chapter 4

Results

Not only quantitative data through the questionnaire, but also qualitative data through the interviews were collected for addressing to the research problem of the study. In this part, the results of the study were presented in two general titles as “questionnaire results” and “interview results” by comparing three countries.

4.1. Questionnaire results

Under this title, the results obtained through 170 participants were presented comparatively in terms of countries.

4.1.1. Typical class size. Typical class sizes mentioned by the teachers in different countries were summarized in Table 2.

Table 2

Typical Class Sizes in Three Countries

Number of the Students	<u>Turkey</u>		<u>Italy</u>		<u>Finland</u>	
	f (n)	p (%)	f (n)	p (%)	f (n)	p (%)
Less than 20 students	18	20.22	12	30	21	51.22
20 to 29 students	52	58.42	28	20	20	48.78
29 to 35 students	18	20.22	-	-	-	-
More than 35 students	1	1.12	-	-	-	-

Regarding the typical class size, of Turkish participants, 18 teachers indicated a typical class size as less than 20 students, 21 teachers as 20 to 29 students, 18 teachers as 29 to 35 students and 1 teacher as more than 35 students. Of the Italian participants, 12 teachers indicated a typical class size as less than 20 students and 28 teachers as 20 to 29 students. Of

the Finnish participants, 21 indicated typical class size as less than 20 students and 20 teachers as 20 to 29 students.

4.1.2. The reasons for carrying out the assessment. Of all participants, 162 expressed that they assessed their students at 4th grade, but remaining 8 did not assess their students at 4th grade. The teachers were asked why the assessment was carried out in their classrooms. They were given a list of possible reasons for classroom assessment and allowed to select more than one reason. Some teachers reported only one reason, but some others reported more than one. The reasons for assessment varied across countries. The frequency of reasons to carry out assessment was given in Table 3.

Table 3

The Reasons for Assessment

The reasons for Assessment	<u>Turkey</u>		<u>Italy</u>		<u>Finland</u>	
	f (n)	p (%)	f (n)	p (%)	f (n)	p (%)
To check students strengths and weaknesses for lesson plan	59	66.29	22	55	29	70.73
To provide information to parents and students	44	49.43	20	50	39	95.12
To place students	10	11.23	7	17.5	2	4.87
To motivate students to learn	47	52.81	17	42.5	30	73.17
To provide a record of students' achievement for success and failure	68	76.40	18	45	19	46.34

Of Turkish teachers, 68 expressed to implement assessment for providing a record of students' achievement for success and failure, 59 for checking students' strengths and

weaknesses for lesson plan, 47 for motivating students to learn, 44 for providing information to parents and students and 10 for placing the students.

Of Italian teachers, 22 expressed to implement assessment for checking students' strengths and weaknesses for lesson plan, 20 for providing information to parents and students, 18 for providing a record of students' achievement for success and failure, 17 for motivating students and 7 for placing the students.

Of Finnish teachers, 39 expressed to implement assessment for providing information to parents and students, 30 for motivating students to learn, 29 for checking students' strengths and weaknesses for lesson plan, 19 for providing a record of students' achievement for success and failure, and 2 for placing the students.

4.1.3. Responsible body for the preparation of the assessment. The teachers were asked who was responsible for the preparation of the assessment. They were given a list of responsible body for the preparation of assessment and allowed to select more than one body (if any). Some teachers reported only one body, but some others reported more than one. Their responses about the responsible body for the preparation of assessment were given in Table 4.

Table 4

Responsible Body for the Preparation of Assessment

Responsible Body	<u>Turkey</u>		<u>Italy</u>		<u>Finland</u>	
	f (n)	p (%)	f (n)	p (%)	f (n)	p (%)
A foreign language teacher	59	66.29	15	37.5	34	82.92
A primary teacher	12	13.48	12	30	10	24.39
A group of language teachers within the school	29	32.58	19	47.5	8	19.51
Local education authority	3	3.37	-	-	7	17.07
National education authority	9	10.11	4	17.07	7	17.07

Of Turkish teachers, 59 indicated a foreign teacher, 12 indicated a primary teacher, 29 indicated a group of language teachers within the school, 3 indicated local education authority and 9 indicated national education authority as to be responsible body for the preparation of assessment.

Of Italian teachers, 15 indicated a foreign language teacher, 12 indicated a primary teacher, 19 indicated a group of language teachers within the school and 4 indicated national education authority as the responsible body for the preparation of assessment.

Of Finnish teachers, 34 indicated a foreign teacher, 10 indicated a primary teacher, 8 indicated a group of language teachers within the school, 7 indicated local education authority and 7 indicated national education authority as to be responsible body for the preparation of assessment.

4.1.4. Reporting the exam results. The teachers were asked the ways to report the exam results and select among three alternatives. Table 5 summarized the ways of reporting exam results.

Table 5

The Ways of Reporting Exam Results

Reporting the Exam Results						
	<u>Turkey</u>		<u>Italy</u>		<u>Finland</u>	
	f (n)	p (%)	f (n)	p (%)	f (n)	p (%)
Numerical grades	72	80.89	26	65	18	43.90
Comments	4	4.49	5	12.5	8	19.51
Grade and comments	22	24.72	11	27.5	31	75.61

Of Turkish teachers, 72 used numeral grades, 4 used comments and 22 used grade and comments for reporting the exam results. 83 teachers reported that they informed their students about the criteria of their assessment, but 6 did not.

Of Italian teachers, 26 used numeral grades, 5 used comments and 22 used grade and comments for reporting the exam results. 38 teachers reported that they informed their students about the criteria of their assessment, but 2 did not.

Of Finnish teachers, 18 used numeral grades, 8 used comments and 31 used grade and comments for reporting the exam results. 35 teachers reported that they informed their students about the criteria of their assessment, but 6 did not.

4.1.5. Additional training for the assessment in primary schools. The teachers were asked to be additionally trained to prepare and perform assessment of students at primary schools and allowed to select among the alternatives (more than one if any). Most of the teachers reported that they received additional education for the assessment at primary schools. Some teachers reported only one educational type, but some others reported more than one. The types of education they received were given in Table 6.

Table 6

The Types of Education They Received for Assessment in Primary Schools

The Types of Education	<u>Turkey</u>		<u>Italy</u>		<u>Finland</u>	
	f (n)	p (%)	f (n)	p (%)	f (n)	p (%)
At university	85	95.50	31	77.5	36	87.81
In seminars organized by the Ministry of Education	2	2.24	5	12.5	6	14.63
From colleagues	10	11.23	3	7.5	12	29.26
In ELT seminars and conferences	20	22.47	8	20	5	12.19

Of Turkish teachers, most ($f = 85$) received additional training at university, 20 in ELT seminars and conferences, 10 from colleagues and 2 in seminars organized by the Ministry of Education for preparing and performing assessment of students at primary schools.

Of Italian teachers, all received additional training for preparing and performing assessment of students at primary schools. 31 received this additional training at university, 5 received in seminars organized by the Ministry of Education, 3 received from colleagues and 8 received in ELT seminars and conferences.

Of Finnish teachers, all received additional training for preparing and performing assessment of students at primary schools. 36 received this additional training at university, 6 received in seminars organized by the Ministry of Education, 12 received from colleagues and 5 received in ELT seminars and conferences.

4.1.6. The types of assessment used in FL classroom. The teachers were asked to indicate what types of assessment they used in 4th grade FL classroom. They were given types of assessment used in classes and required to report the frequency of using any of these types of assessment in 4th grade FL class. Teachers' responses to the items are given in Table 7, Table 8 and Table 9 considering their countries.

Based upon mean average of each item, the most frequently used assessment type was “*my own –made paper and pencil test*” ($\bar{X} = 4.03$) whereas the least frequently used assessment type was “*peer assessment*” ($\bar{X} = 2.29$) among Turkish teachers. Other assessment types used respectively were oral interview ($\bar{X} = 2.78$), standardized paper and pencil test from the national or local EFL education authority ($\bar{X} = 2.65$), performance-based test ($\bar{X} = 2.65$), self-assessment ($\bar{X} = 2.64$), structured observation leading to a written description of students' performance ($\bar{X} = 2.60$) and language portfolio ($\bar{X} = 2.47$).

The most frequently used assessment type was “*oral interview*” ($\bar{X} = 4.27$) whereas the least frequently used assessment type was “*language portfolio*” ($\bar{X} = 2.82$) among Italian

teachers. Other assessment types used respectively were my own-made paper and pencil test ($\bar{X} = 3.8$), performance-based test ($\bar{X} = 3.37$), standardized paper and pencil test from the national or local EFL education authority ($\bar{X} = 3.07$), structured observation leading to a written description of students' performance ($\bar{X} = 3.07$), peer-assessment ($\bar{X} = 2.97$) and language portfolio ($\bar{X} = 2.82$)

The most frequently used assessment type was "*my own –made paper and pencil test*" ($\bar{X} = 3.9$) whereas the least frequently used assessment types were "*peer-assessment*" ($\bar{X} = 2.07$) and "*Standardized paper and pencil test from the national or local EFL education authority*" ($\bar{X} = 2.07$) among Finnish teachers. Other assessment types used respectively were self-assessment ($\bar{X} = 3.6$), oral interview ($\bar{X} = 3.04$), performance-based assessment ($\bar{X} = 2.6$), language portfolio ($\bar{X} = 2.34$) and structured observation leading to a written description of students' performance ($\bar{X} = 2.29$).

Table 7

Frequency of Types of Assessment Used in FL Classroom (Turkey)

Frequency of Types of Assessment Used in FL Classroom (TURKEY)												
<u>Items</u>	\bar{X}	<u>Always</u>		<u>Frequently</u>		<u>Sometimes</u>		<u>Rarely</u>		<u>Never</u>		
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	
My own-made paper and pencil test	4.03	35	39.3	28	31.5	21	23.6	4	4.5	1	1.1	
Standardized paper and pencil test from national or local EFL education authority	2.65	9	10.1	15	16.9	28	31.5	10	11.2	27	30.3	
Performance-based test	2.65	8	9	14	15.7	26	29.2	21	23.6	20	22.5	
Structured observation leading to a written description of students' performance	2.60	7	7.9	14	15.7	25	28.1	23	25.8	20	22.5	
Oral interview	2.78	6	6.7	17	19.1	33	37.1	18	20.2	15	16.9	
Language portfolio	2.47	4	4.5	17	19.1	23	25.8	18	20.2	27	30.3	
Self-assessment	2.64	5	5.6	17	19.1	29	32.6	17	19.1	21	23.6	
Peer-assessment	2.29	2	2.2	10	11.2	28	31.5	21	23.6	28	31.5	

Table 8

Frequency of Types of Assessment Used in FL Classroom (Italy)

Frequency of Types of Assessment Used in FL Classroom (ITALY)												
<u>Items</u>	\bar{X}	<u>Always</u>		<u>Frequently</u>		<u>Sometimes</u>		<u>Rarely</u>		<u>Never</u>		
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	
My own-made paper and pencil test	3.8	9	22.5	21	52.5	3	7.5	7	17.5	-	-	
Standardized paper and pencil test from national or local EFL education authority	3.07	2	5	17	42.5	9	22.5	6	15	6	15	
Performance-based test	3.37	4	10	17	42.5	11	27.5	6	15	2	5	
Structured observation leading to a written description of students' performance	3.07	4	10	11	27.5	12	30	10	25	3	7.5	
Oral interview	4.27	18	45	17	42.5	3	7.5	2	5	-	-	
Language portfolio	2.82	2	5	12	30	13	32.5	3	7.5	10	25	
Self-assessment	3.02	5	12.5	8	20	16	40	5	12.5	6	15	
Peer-assessment	2.97	5	12.5	7	17.5	16	40	6	15	6	15	

Table 9

Frequency of Types of Assessment Used in FL Classroom (Finland)

Frequency of Types of Assessment Used in FL Classroom (FINLAND)												
<u>Items</u>	\bar{X}	<u>Always</u>		<u>Frequently</u>		<u>Sometimes</u>		<u>Rarely</u>		<u>Never</u>		
		f(n)	P(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	
My own-made paper and pencil test	3.9	5	12.2	27	65.9	9	22	-	-	-	-	
Standardized paper and pencil test from national or local EFL education authority	2.07	1	2.4	6	14.6	6	14.6	10	24.4	18	43.9	
Performance-based test	2.6	-	-	7	17.1	15	36.6	15	36.6	4	9.8	
Structured observation leading to a written description of students' performance	2.29	-	-	6	14.6	11	26.8	13	31.7	11	26.8	
Oral interview	3.04	2	4.9	5	12.2	29	70.7	3	7.3	2	4.9	
Language portfolio	2.34	1	2.4	5	12.2	14	34.1	8	19.5	13	31.7	
Self-assessment	3.6	5	12.2	20	48.8	12	29.3	3	7.3	1	2.4	
Peer-assessment	2.07	1	2.4	7	17.1	17	41.5	12	29.3	4	9.8	

In order to examine whether the types of assessment teachers used in 4th grade FL classes varied according to their country, one way analysis of variance (ANOVA) was performed.

Table 10

ANOVA Results in Terms of the Types of Assessment Teachers Used

ANOVA Results in Terms of the Types of Assessment Teachers Used					
Country	f(n)	\bar{X}	SD	F test	Comparison
Turkey	89	22.13	6.28	F (2, 167) = 8.412, $p < 0.05$	Italy > Turkey
Italy	40	26.42	5.32		Italy > Finland
Finland	41	22.60	4.13		

As presented in Table 10, the results of ANOVA showed that teachers' frequency of using various assessment types differed significantly in terms of country [F (2, 167) = 8.412, $p < 0.05$]. Follow up analysis using a post hoc comparison with Tukey test was conducted to observe pair-wise differences among the countries. The post hoc analysis showed that Italian teachers' frequency of using various assessment types ($\bar{X} = 26.42$, SD = 5.32) was significantly higher than Finnish teachers ($\bar{X} = 22.60$, SD = 4.13) and Turkish teachers ($\bar{X} = 22.13$, SD = 6.28)

4.1.7. Content / skills – focus in assessment. The teachers were asked what kind of content and skills they assessed in the exams of 4th grades. They were given a list of content and / or skills and required to report how frequently they assessed any of these skills in the exams. Teachers' responses to the items are given in Table 11, Table 12 and Table 13 considering their countries.

For Turkish teachers, based upon the mean average of the items, the most frequently assessed content / skill at 4th grade exams was “*lexis (Vocabulary)*” ($\bar{X} = 4.46$). Following this

most frequent one, other frequently used content / skills were reading skills, grammar and writing skills. The least frequently assessed content / skill was “*pronunciation*” ($\bar{X} = 2.57$)

For Italian teachers, based upon the mean average of the items, the most frequently assessed content / skill at 4th grade exams was “*lexis (Vocabulary)*” ($\bar{X} = 4.35$). Following this most frequent one, other frequently used content / skills were speaking skills, listening skills, reading skills, writing skills, grammar and pronunciation respectively.

As for Finnish teachers, based upon the mean average of the items, the most frequently assessed content / skill at 4th grade exams was “*lexis (Vocabulary)*” ($\bar{X} = 4.6$). Following this most frequent one, other frequently used content / skills were grammar, listening skills, reading skills, writing skills and spelling. The content / skills of spelling and speaking skills were comparatively less used.

Table 11

Frequency of Content / Skills – Focus in Assessment (Turkey)

Frequency of Content / Skills – Focus in Assessment (TURKEY)											
<u>Items</u>	\bar{X}	<u>Always</u>		<u>Frequently</u>		<u>Sometimes</u>		<u>Rarely</u>		<u>Never</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Grammar	3.57	22	24.7	29	32.6	19	21.3	16	18	3	3.4
Lexis(Vocabulary)	4.46	46	51.7	38	42.7	5	5.6	-	-	-	-
Spelling	2.85	6	6.7	22	24.7	26	29.2	23	25.8	12	13.5
Speaking skills	2.77	8	9	18	20.2	23	25.8	26	29.2	14	15.7
Listening Skills	2.59	8	9	16	18	19	21.3	24	27	22	24.7
Reading skills	3.64	21	23.6	33	37.1	21	23.6	10	11.2	4	4.5
Writing skills	3.26	17	19.1	23	25.8	24	27	17	19.1	8	9
Pronunciation	2.57	4	4.5	21	23.6	20	22.5	21	23.6	23	25.8

Table 12

Frequency of Content / Skills – Focus in Assessment (Italy)

Frequency of Content / Skills – Focus in Assessment (ITALY)											
<u>Items</u>	\bar{X}	<u>Always</u>		<u>Frequently</u>		<u>Sometimes</u>		<u>Rarely</u>		<u>Never</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Grammar	4	13	32.5	16	40	9	22.5	2	5	-	-
Lexis(Vocabulary)	4.35	17	42.5	20	50	3	7.5	-	-	-	-
Spelling	3.82	11	27.5	14	35	12	30	3	7.5	-	-
Speaking skills	4.32	21	52.5	11	27.5	8	20	-	-	-	-
Listening Skills	4.25	18	45	14	35	8	20	-	-	-	-
Reading skills	4.25	18	45	14	35	8	20	-	-	-	-
Writing skills	4.05	16	40	12	20	10	25	2	5	-	-
Pronunciation	3.95	15	37.5	11	27.5	12	29.3	1	2.5	1	2.5

Table 13

Frequency of Content / Skills – Focus in Assessment (Finland)

Frequency of Content / Skills – Focus in Assessment (FINLAND)											
<u>Items</u>	\bar{X}	<u>Always</u>		<u>Frequently</u>		<u>Sometimes</u>		<u>Rarely</u>		<u>Never</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Grammar	4.39	20	48.8	17	41.5	4	9.8	-	-	4.39	20
Lexis(Vocabulary)	4.6	27	65.9	12	29.3	2	4.9	-	-	4.6	27
Spelling	3.9	11	26.8	18	43.9	9	22	3	7.3	3.9	11
Speaking skills	3.39	4	9.8	12	29.3	21	51.2	4	9.8	3.39	4
Listening Skills	4.24	15	36.6	21	51.2	5	12.2	-	-	4.24	15
Reading skills	4.07	12	29.3	21	51.2	7	17.1	1	2.4	4.07	12
Writing skills	4.07	14	34.1	16	39	11	26.8	-	-	4.07	14
Pronunciation	2.97	2	4.9	7	17.1	20	48.8	12	29.3	2.97	2

In order to examine whether the types of content / skills teachers assessed at 4th grade language exams changed according to their country, one way analysis of variance (ANOVA) was performed. ANOVA results are given in Table 14.

Table 14

ANOVA Results in Terms of Types of Content / Skills Teachers Assessed at 4th Grade Language Exams

ANOVA Results in Terms of the Types of Content / Skills Teachers Assessed at 4th Grade Language Exams					
Country	f(n)	\bar{X}	SD	F test	Comparison
Turkey	89	25.74	5.53	F (2, 167) = 37.342, $p < 0.05$	Italy > Turkey
Italy	40	33.0	4.65		Finland > Turkey
Finland	41	31.65	4.02		

As presented in Table 14, the results of ANOVA showed that frequency of content / skills teachers assessed at 4th grade language exams differed significantly in terms of country [F (2, 167) = 37.342, $p < 0.05$]. Follow up analysis using post hoc comparison with Tukey test was conducted to observe pair-wise differences among the countries. The post hoc analysis showed that the frequency of content / skills that Italian teachers ($\bar{X} = 33.0$, SD = 4.65) and Finnish teachers ($\bar{X} = 31.65$, SD = 4.02) were significantly higher than that of Turkish teachers ($\bar{X} = 25.74$, SD = 5.53).

4.1.8. Kinds of exercises used in FL classroom. The teachers were asked what kinds of exercises they used in 4th grade FL classroom. They were given a list of exercises used in FL classes and required to report how frequently they used any of these exercises in their classroom. Teachers' responses to the items are given in Table 15, Table 16 and Table 17 considering their countries.

Table 15

Frequency of Kinds of Exercises Used in FL Classroom (Turkey)

Frequency of Kinds of Exercises Used in FL Classroom (TURKEY)											
<u>Items</u>	X	<u>Always</u>		<u>Frequently</u>		<u>Sometimes</u>		<u>Rarely</u>		<u>Never</u>	
		f(n)	p(%)	f(n)	p(%)	f (n)	p(%)	f (n)	p(%)	f (n)	p(%)
Filing gaps	4.42	46	51.7	36	40.4	6	6.7	1	1.1	-	-
Matching	4.55	52	58.4	34	38.2	3	3.4	-	-	-	-
Role playing	2.34	4	4.5	11	12.4	22	24.7	27	30.3	25	28.1
Writing words, short texts	3.35	17	19.1	23	25.8	29	32.6	15	16.9	5	5.6
Comprehension Qs about a short text	3.93	27	30.3	39	43.8	15	16.9	6	6.7	2	2.2
Dictation	2.05	1	1.1	10	11.2	19	21.3	22	24.7	37	41.6
Grammatical activities	3.55	23	25.8	29	32.6	16	18	16	18	5	5.6
Students create their own sentences/texts	2.70	5	5.6	18	20.2	27	30.3	24	27	15	16.9
Listening to audio material	2.23	6	6.7	11	12.4	16	18	21	23.6	35	39.3
Making a dialogue	2.53	5	5.6	12	13.5	28	31.5	25	28.1	19	21.3
Students repeat vocabulary or sentences	2.56	11	12.4	11	12.4	23	25.8	16	18	28	31.5
Multiple Choice	3.84	27	30.3	31	34.8	22	24.7	8	9	1	1.1
Reading words and sentences	3.01	13	14.6	23	25.8	18	20.2	22	24.7	13	14.6
Oral description of a picture	2.24	3	3.4	16	18	12	13.5	27	30.3	31	34.8
Spelling	2.25	4	4.5	8	9	24	27	24	27	29	32.6

Table 16

Frequency of Kinds of Exercises Used in FL Classroom (Italy)

Frequency of Kinds of Exercises Used in FL Classroom (ITALY)											
<u>Items</u>	X	<u>Always</u>		<u>Frequently</u>		<u>Sometimes</u>		<u>Rarely</u>		<u>Never</u>	
		f(n)	p(%)	f(n)	p(%)	f (n)	p(%)	f (n)	p(%)	f (n)	p(%)
Filing gaps	4.12	12	30	22	55	5	12.5	1	2.5	-	-
Matching	4.12	12	30	22	55	5	12.5	1	2.5	-	-
Role playing	3.35	9	22.5	13	32.5	5	12.5	9	22.5	4	20
Writing words, short texts	3.9	9	22.5	22	55	6	15	2	5	1	2.5
Comprehension Qs about a short text	4.3	18	45	17	42.5	4	20	1	2.5	-	-
Dictation	2.17	3	7.5	4	20	5	12.5	13	32.5	15	37.5
Grammatical activities	3.7	7	17.5	18	45	12	30	2	5	1	2.5
Students create their own sentences/texts	2.77	3	7.5	9	22.5	13	32.5	6	15	9	22.5
Listening to audio material	4.1	17	42.5	13	32.5	7	17.5	3	7.5	-	-
Making a dialogue	3.47	7	17.5	13	32.5	15	37.5	2	5	3	7.5
Students repeat vocabulary or sentences	3.32	8	20	13	32.5	6	16	10	25	3	7.5
Multiple Choice	3.92	6	15	26	65	7	17.5	1	2.5	-	-
Reading words and sentences	3.87	11	27.5	19	47.5	4	10	6	15	-	-
Oral description of a picture	3.55	8	20	14	35	13	32.5	2	5	3	7.5
Spelling	2.9	3	7.5	15	37.5	3	7.5	13	32.5	6	15

Table 17

Frequency of Kinds of Exercises Used in FL Classroom (Finland)

Frequency of Kinds of Exercises Used in FL Classroom (FINLAND)											
<u>Items</u>	X	<u>Always</u>		<u>Frequently</u>		<u>Sometimes</u>		<u>Rarely</u>		<u>Never</u>	
		f(n)	p(%)	f(n)	p(%)	f (n)	p(%)	f (n)	p(%)	f (n)	p(%)
Filing gaps	4.07	10	24.4	24	58.5	7	17.1	-	-	-	-
Matching	3.92	7	17.1	26	63.4	6	14.6	2	4.9	-	-
Role playing	2.24	-	-	-	-	18	43.9	15	36.6	8	19.5
Writing words, short texts	4.21	16	39	19	46.3	5	12.2	1	2.4	-	-
Comprehension Qs about a short text	3.75	9	22	18	43.9	11	26.8	1	2.4	2	4.9
Dictation	2.48	-	-	6	14.6	13	31.7	17	41.5	5	12.2
Grammatical activities	4.02	12	29.3	20	48.8	7	17.1	2	4.9	-	-
Students create their own sentences/texts	3.73	7	17.1	20	48.8	10	24.4	4	9.8	-	-
Listening to audio material	3.9	7	17.1	25	61	8	19.5	-	-	1	2.4
Making a dialogue	2.65	-	-	4	9.8	22	53.7	12	29.3	3	7.3
Students repeat vocabulary or sentences	1.85	-	-	4	9.8	5	12.2	13	31.7	19	46.3
Multiple Choice	3.07	2	4.9	11	26.8	18	43.9	8	19.5	2	4.9
Reading words and sentences	2.65	3	7.3	8	19.5	7	17.1	18	43.9	5	12.2
Oral description of a picture	2.43	2	4.9	4	9.8	11	26.8	17	41.5	7	17.1
Spelling	2.68	4	9.8	5	12.2	12	29.3	14	34.1	6	14.6

For Turkish teachers, based upon the mean average of the items, the most frequently used exercise in FL classroom was “*matching*” ($\bar{X} = 4.55$) Following this most frequent one, other frequently used exercises were filling gaps ($\bar{X} = 4.42$), comprehension questions about a short text ($\bar{X} = 3.93$), multiple choice ($\bar{X} = 3.84$) and grammatical activities ($\bar{X} = 3.55$). The least frequently used exercise in FL classroom was “*dictation*” ($\bar{X} = 2.05$). Other less used exercises were listening to audio material ($\bar{X} = 2.23$), oral description of a picture ($\bar{X} = 2.24$), spelling ($\bar{X} = 2.25$) and role playing ($\bar{X} = 2.34$).

For Italian teachers, based upon the mean average of the items, the most frequently used exercise in FL classroom was “*comprehension questions about a short text*” ($\bar{X} = 4.3$) Following this most frequent one, other frequently used exercises were filling gaps ($\bar{X} = 4.12$), matching ($\bar{X} = 4.12$), listening to audio material ($\bar{X} = 4.1$), multiple choice ($\bar{X} = 3.92$), writing words, short texts ($\bar{X} = 3.9$), reading words and sentences ($\bar{X} = 3.87$), grammatical activities ($\bar{X} = 3.7$) and oral description of a picture ($\bar{X} = 3.55$). The least frequently used exercise in FL classroom was “*dictation*” ($\bar{X} = 2.17$).

As for Finnish teachers, based upon the mean average of the items, the most frequently used exercise in FL classroom was “*writing words, short texts*” ($\bar{X} = 4.21$). Following this most frequent one, other frequently used exercises were filling gaps ($\bar{X} = 4.07$), grammatical activities ($\bar{X} = 4.02$), matching ($\bar{X} = 3.92$), comprehension questions about a short text ($\bar{X} = 3.75$) and students create their own sentences / texts ($\bar{X} = 3.73$). The least frequently used exercise in FL classroom was “*students repeat vocabulary or sentences*” ($\bar{X} = 1.85$). Other less used exercises were role playing ($\bar{X} = 2.24$), oral description of a picture ($\bar{X} = 2.43$), dictation ($\bar{X} = 2.48$), making a dialogue ($\bar{X} = 2.65$) and reading words and sentences ($\bar{X} = 2.65$).

In order to examine whether the kinds of activities teachers used at 4th grade language classroom changed according to their country, one way analysis of variance (ANOVA) was performed. ANOVA results are given in Table 18

Table 18

ANOVA Results in Terms of Kinds of Exercises Used in FL Classroom

ANOVA Results in Terms of Kinds of Exercises Used in FL Classroom					
Country	f(n)	\bar{X}	SD	F test	Comparison
Turkey	89	45.62	9.78	F (2, 167) = 10.846, $p < 0.05$	Italy > Turkey
Italy	40	53,6	9.52		Italy > Finland
Finland	41	47.73	6.24		

As presented in Table 18, the results of ANOVA showed that frequency of types of exercises teachers used in FL classroom differed significantly in terms of country [F (2, 167) = 10.846, $p < 0.05$]. Follow up analysis using post hoc comparison with Tukey test was conducted to observe pair-wise differences among the countries. The post hoc analysis showed that the frequency of kinds of exercises used by Italian teachers in FL classroom ($\bar{X} = 53.6$, SD = 9.52) was significantly higher than that of Finnish teachers ($\bar{X} = 47.73$ SD = 6.24) and Turkish teachers ($\bar{X} = 45.62$, SD = 9.78).

4.1.9. Beliefs about the types of assessment. The teachers were asked to indicate their beliefs about the types of assessment to be used for 4th grade students. The responses to beliefs items about the types of assessment are presented in Table 19, Table 20 and Table 21.

Table 19

Beliefs about the Types of Assessment (Turkey)

Beliefs about the Types of Assessment (TURKEY)

<u>Items</u>	X	<u>I fully agree</u>		<u>I agree</u>		<u>I partially agree</u>		<u>I don't agree</u>		<u>I don't agree at all</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
My own-made paper and pencil test	3.87	28	31.5	34	38.2	17	19.1	8	9	2	2.2
Standardized paper and pencil test from the national or local EFL education authority	3.10	6	6.7	26	29.2	35	39.3	15	16.9	7	7.9
Performance-based test	4.04	30	33.7	39	43.8	15	16.9	4	4.5	1	1.1
Structured observation leading to a written description of students' performance	3.43	13	14.6	31	34.8	31	34.8	10	11.2	4	4.5
Oral interview	4.02	33	37.1	35	39.3	12	13.5	8	9	1	1.1
Language portfolio	3.94	25	28.1	38	42.7	22	24.7	4	4.5	-	-
Self-assessment	3.64	20	22.5	32	36	26	29.2	7	7.9	4	4.5
Peer-assessment	3.50	19	21.3	27	30.3	27	30.3	12	13.5	4	4.5

Table 20

Beliefs about the Types of Assessment (Italy)

Beliefs about the Types of Assessment (ITALY)

<u>Items</u>	X	<u>I fully agree</u>		<u>I agree</u>		<u>I partially agree</u>		<u>I don't agree</u>		<u>I don't agree at all</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
My own-made paper and pencil test	3.8	9	22.5	17	42.5	11	27.5	3	7.5	-	-
Standardized paper and pencil test from the national or local EFL education authority	3.52	6	15	18	45	10	25	3	7.5	3	7.5
Performance-based test	3.85	11	27.5	16	40	11	27.5	-	-	2	5
Structured observation leading to a written description of students' performance	3.85	7	17.5	22	55	9	22.5	2	5	-	-
Oral interview	4.42	20	50	19	47.5	-	-	-	-	1	2.5
Language portfolio	3.2	7	17.5	12	30	11	27.5	2	5	8	20
Self-assessment	3.75	9	22.5	19	47.5	8	20	1	2.5	3	7.5
Peer-assessment	3.55	7	17.5	18	45	8	20	4	10	3	7.5

Table 21

Beliefs about the Types of Assessment (Finland)

Beliefs about the Types of Assessment (FINLAND)

<u>Items</u>	X	<u>I fully agree</u>		<u>I agree</u>		<u>I partially agree</u>		<u>I don't agree</u>		<u>I don't agree at all</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
My own-made paper and pencil test	4.26	17	41.5	18	43.9	6	14.6	-	-	-	-
Standardized paper and pencil test from the national or local EFL education authority	2.65	3	7.3	6	14.6	12	29.3	14	34.1	6	14.6
Performance-based test	3.48	4	9.8	16	39	17	41.5	4	9.8	-	-
Structured observation leading to a written description of students' performance	3.36	4	9.8	16	39	12	29.3	9	22	-	-
Oral interview	4.02	11	26.8	21	51.2	8	19.5	1	2.4	-	-
Language portfolio	3.75	6	14.6	22	53.7	10	24.4	3	7.3	-	-
Self-assessment	4.26	18	43.9	17	41.5	5	12.2	1	2.4	-	-
Peer-assessment	3.63	5	12.2	20	48.8	12	29.3	4	9.8	-	-

For Turkish teachers, based upon the mean average of the items, they agreed that the types of assessment of performance based test ($\bar{X} = 4.04$) and oral interview ($\bar{X} = 4.02$) should be used for 4th graders. They tended to agree that language portfolio ($\bar{X} = 3.94$), teacher own – made paper and pencil test ($\bar{X} = 3.87$), self-assessment ($\bar{X} = 3.64$), peer assessment ($\bar{X} = 3.50$), structured observation ($\bar{X} = 3.43$) and standardize paper and pencil test ($\bar{X} = 3.10$) should be used for 4th grade students.

For Italian teachers, based upon the mean average of the items, they agreed that the types of assessment of oral interview ($\bar{X} = 4.42$) should be used for 4th grade students. They tended to agree that performance-based test ($\bar{X} = 3.85$), structured observation ($\bar{X} = 3.85$), teacher own-made paper and paper and pencil test ($\bar{X} = 3.8$), self-assessment ($\bar{X} = 3.75$), peer assessment ($\bar{X} = 3.55$) and standardized paper and pencil test ($\bar{X} = 3.52$) should be used for 4th grade students.

As for Finnish teachers, based upon the mean average of the items, they agreed that the types of assessment of self-assessment ($\bar{X} = 4.26$), teacher own-made paper and pencil test ($\bar{X} = 4.26$) and oral interview ($\bar{X} = 4.02$) should be used for 4th grade students. They tended to agree that language portfolio ($\bar{X} = 3.75$) and peer-assessment ($\bar{X} = 3.63$) should be used for 4th grade students.

In order to examine whether teachers' beliefs about using the types of assessment for 4th graders changed according to their country, one way analysis of variance (ANOVA) was performed. ANOVA results are given in Table 22.

Table 22

ANOVA Results in Terms of Teachers' Beliefs about the Types of Assessment to Be Used in FL Classroom

ANOVA Results in Terms of Teachers' Beliefs about the Types of Assessment to Be Used in FL Classroom					
Country	f(n)	\bar{X}	SD	F test	Comparison
Turkey	89	29.57	4.31	F (2, 167) = 0.161 , p = 0.852	-
Italy	40	29.95	4.69		-
Finland	41	29.46	3.03		

As presented in Table 22, the results of ANOVA showed that teachers' beliefs about types of assessment to be used at 4th grade FL classroom did not differ significantly in terms of country [F (2, 167) = 0.161 , p = 0.852].

4.1.10. Beliefs about the content / skill – focus in assessment. Teachers were asked to indicate their beliefs about the content / skills – focus needed to be assessed for the 4th grade students. The responses to beliefs items about content / skills – focus are given in Table 23, Table 24 and Table 25.

For Turkish teachers, based upon the mean average of the items, they tended to fully agree that the content / skills – focus needed to be assessed for 4th graders were speaking skills (\bar{X} = 4.56) and listening skills (\bar{X} = 4.51). They agreed that the content / skills – focus needed to be assessed were Lexis (vocabulary) (\bar{X} = 4.41), reading skills (\bar{X} = 4.26), pronunciation (\bar{X} = 4.14). They tended to agree on writing skills (\bar{X} = 3.82) and spelling (\bar{X} = 3.75) and they tended to partially agree on grammar (\bar{X} = 2.96) as to be content / skill – focuses needed to be assessed for 4th graders.

For Italian teachers, based upon the mean average of the items, they tended to fully agree that the content / skills – focus needed to be assessed for 4th graders were speaking

skills ($\bar{X} = 4.67$), listening skills ($\bar{X} = 4.67$) and reading skills ($\bar{X} = 4.52$). They agreed that the content / skills – focus needed to be assessed were lexis (vocabulary) ($\bar{X} = 4.4$), pronunciation ($\bar{X} = 4.2$) and writing skills ($\bar{X} = 4.17$). They tended to agree on spelling ($\bar{X} = 3.85$) and grammar ($\bar{X} = 3.7$) as to be content / skill – focuses needed to be assessed for 4th graders.

As for Finnish teachers, based upon the mean average of the items, they tended to fully agree that the content / skills – focus needed to be assessed for 4th graders were Lexis (vocabulary) ($\bar{X} = 4.60$). They agreed that the content / skills – focus needed to be assessed were listening skills ($\bar{X} = 4.41$), speaking skills ($\bar{X} = 4.36$), reading skills ($\bar{X} = 4.19$) and writing skills ($\bar{X} = 3.97$). They tended to agree on pronunciation ($\bar{X} = 3.8$), grammar ($\bar{X} = 3.78$) and spelling ($\bar{X} = 3.51$) as to be content / skill – focuses needed to be assessed for 4th graders.

Table 23

Beliefs about the Content / Skill – Focus in Assessment (Turkey)

Beliefs about the Content / Skill – Focus in Assessment (TURKEY)											
<u>Items</u>	\bar{X}	<u>FA</u>		<u>A</u>		<u>PA</u>		<u>DA</u>		<u>DAA</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Grammar	2.96	7	7.9	21	23.6	34	38.2	16	18	11	12.4
Lexis(Vocabulary)	4.41	45	50.6	36	40.4	8	9	-	-	-	-
Spelling	3.75	26	29.2	27	30.3	25	28.1	10	11.2	1	1.1
Speaking skills	4.59	61	68.5	20	22.5	8	9	-	-	-	-
Listening Skills	4.51	55	61.8	25	28.1	9	10.1	-	-	-	-
Reading skills	4.26	43	48.3	29	32.6	16	18	-	-	1	1.1
Writing skills	3.82	28	31.5	28	31.5	23	25.8	9	10.1	1	1.1
Pronunciation	4.14	40	44.9	28	31.5	15	16.9	6	6.7	-	-

FA: I fully agree, A: I agree, PA: I partially agree, DA: I don't agree, DAA: I don't agree at all

Table 24

Beliefs about the Content / Skill – Focus in Assessment (Italy)

Beliefs about the Content / Skill – Focus in Assessment (ITALY)											
<u>Items</u>	<u>X</u>	<u>FA</u>		<u>A</u>		<u>PA</u>		<u>DA</u>		<u>DAA</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Grammar	3.7	8	20	15	37.5	15	37.5	1	2.5	1	2.5
Lexis(Vocabulary)	4.4	17	42.5	22	55	1	2.5	-	-	-	-
Spelling	3.85	13	32.5	11	27.5	13	32.5	3	7.5	-	-
Speaking skills	4.67	28	70	11	27.5	1	2.5	-	-	-	-
Listening Skills	4.67	29	72.5	9	22.5	2	5	-	-	-	-
Reading skills	4.52	22	55	17	42.5	1	2.5	-	-	-	-
Writing skills	4.17	16	40	16	40	7	17.5	1	2.5	-	-
Pronunciation	4.2	18	45	13	32.5	8	20	1	2.5	-	-

FA: I fully agree, A: I agree, PA: I partially agree, DA: I don't agree, DAA: I don't agree at all

Table 25

Beliefs about the Content / Skill – Focus in Assessment (Finland)

Beliefs about the Content / Skill – Focus in Assessment (FINLAND)											
<u>Items</u>	<u>X</u>	<u>FA</u>		<u>A</u>		<u>PA</u>		<u>DA</u>		<u>DAA</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Grammar	3.78	8	19.5	17	41.5	15	36.6	1	2.4	-	-
Lexis(Vocabulary)	4.60	25	61	16	39	-	-	-	-	-	-
Spelling	3.51	7	17.1	13	31.7	15	36.6	6	14.6	-	-
Speaking skills	4.36	17	41.5	22	53.7	2	4.9	-	-	-	-
Listening Skills	4.41	18	43.9	22	53.7	1	2.4	-	-	-	-
Reading skills	4.19	14	34.1	22	53.7	4	9.8	1	2.4	-	-
Writing skills	3.97	12	29.3	13	31.7	15	36.6	1	2.4	-	-
Pronunciation	3.8	9	22	19	46.3	9	22	4	9.8	-	-

FA: I fully agree, A: I agree, PA: I partially agree, DA: I don't agree, DAA: I don't agree at all

In order to examine whether teachers' beliefs about the content / skills – focus needed to be assessed for 4th graders changed according to their country, one way analysis of variance (ANOVA) was performed. ANOVA results are given in Table 26

Table 26

ANOVA Results in Terms of Teachers' Beliefs about the Content / Skills – Focus Needed to Be Assessed for 4th Graders

ANOVA Results in Terms of Teachers' Beliefs about the Content / Skills – Focus Needed to Be Assessed for 4th Graders					
<u>Country</u>	<u>f(n)</u>	<u>\bar{X}</u>	<u>SD</u>	<u>F test</u>	<u>Comparison</u>
Turkey	89	32.48	4.43	F (2, 167) = 2.585 , p = 0.078	-
Italy	40	34.2	3.4		-
Finland	41	32.56	4.03		

As presented in Table 26, the results of ANOVA showed that teachers' beliefs about the content / skills – focus needed to be assessed for 4th graders did not differ significantly in terms of country [F (2, 167) = 2.585 , p = 0.078]

4.1.11. Beliefs about the kinds of exercises. Teachers were asked to indicate their beliefs about the kinds of exercises needed to be used for 4th grade students. The responses to beliefs about the kinds of exercises needed to be used for 4th grade students are given in Table 27, Table 28 and Table 29.

For Turkish teachers, based upon the mean average of the items, they agreed that the kinds of exercises needed to be used for 4th graders were listening to audio material (\bar{X} = 4.43), matching (\bar{X} = 4.39), oral description of picture (\bar{X} = 4.28), role playing (\bar{X} = 4.26), making a dialogue (\bar{X} = 4.21), comprehension questions about a short text (\bar{X} = 4.17), filling gaps (\bar{X} = 4.17). They tended to agree that students create their own sentences / text (\bar{X} = 3.77), writing words, short texts (\bar{X} = 3.73), reading words and sentences (\bar{X} = 3.69), students

repeat vocabulary or sentences ($\bar{X} = 3.67$), spelling ($\bar{X} = 3.59$) and multiple choice ($\bar{X} = 3.52$) as to be kinds of exercises needed to be used for 4th graders. They partially agree on dictation ($\bar{X} = 3.22$) and grammatical activities ($\bar{X} = 3.03$) needed to be used for 4th graders.

For Italian teachers, based upon the mean average of the items, they agreed that the kinds of exercises needed to be used for 4th graders were listening to audio material ($\bar{X} = 4.42$), comprehension questions about a short test ($\bar{X} = 4.26$), fillings gaps ($\bar{X} = 4.25$), matching ($\bar{X} = 4.25$), oral description of a picture ($\bar{X} = 4.2$), multiple choice ($\bar{X} = 4.15$), role playing ($\bar{X} = 4.1$), writing words, short texts ($\bar{X} = 4.05$), reading words and sentences ($\bar{X} = 3.97$) and making a dialogue ($\bar{X} = 3.97$). They tended to agree on repeating vocabulary or sentences by students ($\bar{X} = 3.72$), spelling ($\bar{X} = 3.62$), students' creation of their own sentences /texts ($\bar{X} = 3.57$), grammatical activities ($\bar{X} = 3.55$), and they partially agreed on dictation as to be kinds of exercises needed to be used for 4th graders.

As for Finnish teachers, based upon the mean average of the items, they agreed that the kinds of exercises needed to be used for 4th graders were listening to audio materials ($\bar{X} = 4.34$), students' creation of their own sentences /texts ($\bar{X} = 4.29$), comprehension questions about a short test ($\bar{X} = 4.26$), matching ($\bar{X} = 4.12$) and writing words, short texts ($\bar{X} = 4.07$), filling gaps ($\bar{X} = 3.97$) and grammatical activities ($\bar{X} = 3.95$). They tended to agree on oral description of a picture ($\bar{X} = 3.75$), reading words and sentences ($\bar{X} = 3.73$), role playing ($\bar{X} = 3.68$) and making a dialogue ($\bar{X} = 3.63$), and they partially agreed on multiple choice ($\bar{X} = 3.36$), dictation ($\bar{X} = 3.31$) and repeating vocabulary or sentences by students ($\bar{X} = 3$).

Table 27

Beliefs about the Kinds of Exercises (Turkey)

<u>Items</u>	<u>X</u>	<u>I fully agree</u>		<u>I agree</u>		<u>I partially agree</u>		<u>I don't agree</u>		<u>I don't agree at all</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Filing gaps	4.17	33	37.1	41	46.1	13	14.6	2	2.2	-	-
Matching	4.39	43	48.3	38	42.7	8	9	-	-	-	-
Role playing	4.26	45	50.6	27	30.3	13	14.6	4	4.5	-	-
Writing words, short texts	3.73	20	22.5	35	39.3	25	28.1	8	9	1	1.1
Comprehension Qs about a short text	4.17	35	39.3	36	40.4	17	19.1	1	1.1	-	-
Dictation	3.22	18	20.2	21	23.6	21	23.6	21	23.6	8	9
Grammatical activities	3.03	7	7.9	19	21.3	38	42.7	20	22.5	5	5.6
Students create their own sentences/texts	3.77	25	28.1	32	36	22	24.7	7	7.9	3	3.4
Listening to audio material	4.43	48	53.9	33	37.1	7	7.9	1	1.1	-	-
Making a dialogue	4.21	44	49.4	26	29.2	15	16.9	2	2.2	2	2.2
Students repeat vocabulary or sentences	3.67	24	27	29	32.6	22	24.7	11	12.4	3	3.4
Multiple Choice	3.52	16	18	35	39.3	21	23.6	14	15.7	3	3.4
Reading words and sentences	3.69	22	24.7	28	31.5	31	34.8	6	6.7	2	2.2
Oral description of a picture	4.28	43	48.3	31	34.8	12	13.5	3	3.4	-	-
Spelling	3.59	20	22.5	32	36	21	23.6	13	14.6	3	3.4

Table 28

Beliefs about the Kinds of Exercises (Italy)

<u>Items</u>	<u>X</u>	<u>I fully agree</u>		<u>I agree</u>		<u>I partially agree</u>		<u>I don't agree</u>		<u>I don't agree at all</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Filing gaps	4.25	13	32.5	24	60	3	7.5	-	-	-	-
Matching	4.25	18	45	21	52.5	1	2.5	-	-	-	-
Role playing	4.1	17	42.5	14	35	7	17.5	-	-	2	5
Writing words, short texts	4.05	13	32.5	16	40	11	27.5	-	-	-	-
Comprehension Qs about a short text	4.4	18	45	20	50	2	5	-	-	-	-
Dictation	3	3	7.5	10	25	15	37.5	8	20	4	10
Grammatical activities	3.55	7	17.5	13	32.5	16	40	3	7.5	1	2.5
Students create their own sentences/texts	3.57	9	22.5	15	37.5	10	25	2	5	4	10
Listening to audio material	4.42	24	60	11	27.5	3	7.5	2	5	-	-
Making a dialogue	3.97	15	37.5	15	37.5	7	17.5	-	-	3	7.5
Students repeat vocabulary or sentences	3.72	10	25	15	37.5	9	22.5	6	15	-	-
Multiple Choice	4.15	14	35	20	50	4	10	2	5	-	-
Reading words and sentences	3.97	12	30	17	42.5	9	22.5	2	5	-	-
Oral description of a picture	4.2	14	35	20	50	6	15	-	-	-	-
Spelling	3.62	10	25	11	27.5	13	32.5	6	15	-	-

Table 29

Beliefs about the Kinds of Exercises (Finland)

<u>Items</u>	<u>X</u>	<u>I fully agree</u>		<u>I agree</u>		<u>I partially agree</u>		<u>I don't agree</u>		<u>I don't agree at all</u>	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Filing gaps	3.97	10	24.4	21	51.2	9	22	1	2.4	-	-
Matching	4.12	10	24.4	26	63.4	5	12.2	-	-	-	-
Role playing	3.68	5	12.2	19	46.3	16	39	1	2.4	-	-
Writing words, short texts	4.07	11	26.8	22	53.7	8	19.5	-	-	-	-
Comprehension Qs about a short text	4.26	15	36.6	22	53.7	4	9.8	-	-	-	-
Dictation	3.31	4	9.8	15	36.6	12	29.3	10	24.4	-	-
Grammatical activities	3.95	10	24.4	19	46.3	12	29.3	-	-	-	-
Students create their own sentences/texts	4.29	16	39	21	51.2	4	9.8	-	-	-	-
Listening to audio material	4.34	16	39	24	58.5	-	-	1	2.4	-	-
Making a dialogue	3.63	8	19.5	14	34.2	15	36.6	4	9.8	-	-
Students repeat vocabulary or sentences	3	5	12.2	9	22	11	26.8	13	31.7	3	7.3
Multiple Choice	3.36	5	12.2	10	24.4	23	56.1	1	2.4	2	4.9
Reading words and sentences	3.73	8	19.5	20	48.8	7	17.1	6	14.6	-	-
Oral description of a picture	3.75	7	17.1	19	46.3	13	31.7	2	4.9	-	-
Spelling	3.43	5	12.2	15	36.6	14	34.1	7	17.1	-	-

In order to examine whether teachers' beliefs about kinds of exercises needed to be used for 4th grade students changed according to their country, one way analysis of variance (ANOVA) was performed. ANOVA results are given in Table 30

Table 30

ANOVA Results in Terms of Teachers' Beliefs about the Kinds of Exercises Needed to Be Used for 4th Grade Students

ANOVA Results in Terms of Teachers' Beliefs about the Kinds of Exercises Needed to Be Used for 4th Grade Students					
<u>Country</u>	<u>N</u>	<u>\bar{X}</u>	<u>SD</u>	<u>F test</u>	<u>Comparison</u>
Turkey	89	58.21	7.74	F (2, 167) = 1.081 , p = 0.342	-
Italy	40	59.42	7.4		-
Finland	41	56.95	7.35		

As presented in Table 30, the results of ANOVA showed that teachers' beliefs about the kinds of exercises needed to be used for 4th grade students did not differ significantly in terms of country [F (2, 167) = 1.081 , p = 0.342].

4.1.12. Teachers' general beliefs about assessment of students' foreign language.

The teachers were asked to indicate their general beliefs about assessment of 4th grade students' foreign language. The responses to the items are given in Table 31, Table 32 and Table 33.

For Turkish teachers, mean scores of their general beliefs of students' foreign language were between $\bar{X} = 4.68$ and $\bar{X} = 2.97$. The item of "*It is difficult to assess speaking skills in the crowded classes*" had the highest mean score whereas the item of "*The results of assessment should be used to place students*" had the lowest mean score. Other items' mean scores were respectively; there should be ready-made exams in the books to assess listening skills ($\bar{X} = 4.19$), students should be assessed to check students strength and weaknesses for

lesson planning ($\bar{X} = 3.98$), parents should always be informed with the results of students' assessment ($\bar{X} = 3.97$), there should be ready-made exams in the books to assess speaking skills ($\bar{X} = 3.93$), students should be assessed by comments, not by numerical grades ($\bar{X} = 3.66$), I need to be more trained in assessing students in primary schools ($\bar{X} = 3.48$), students of a foreign language should be assessed in the primary school ($\bar{X} = 3.26$), students should be assessed to provide a record for success and failure ($\bar{X} = 3.24$), and students should be tested on sentence-level activities ($\bar{X} = 3.15$).

For Italian teachers, mean scores of their general beliefs of students' foreign language were between $\bar{X} = 4.2$ and $\bar{X} = 3.45$. The item of "*Students of a foreign language should be assessed in the primary school*" had highest mean score whereas the item of "*The results of assessment should be used to place students*" had the lowest mean score. Other items' mean scores were respectively; there should be ready-made exams in the books to assess listening skills ($\bar{X} = 4.15$), students should be assessed to check students strength and weaknesses for lesson planning ($\bar{X} = 4.15$), there should be ready-made exams in the books to assess speaking skills ($\bar{X} = 4.15$), parents should always be informed with the results of students' assessment ($\bar{X} = 3.92$), I need to be more trained in assessing students in primary schools ($\bar{X} = 3.72$), It is difficult to assess speaking skills in the crowded classes ($\bar{X} = 3.7$), students should be tested on sentence-level activities ($\bar{X} = 3.7$), students should be assessed to provide a record for success and failure ($\bar{X} = 3.67$) and students should be assessed by comments, not by numerical grades ($\bar{X} = 3.62$).

For Finnish teachers, mean scores of their general beliefs of students' foreign language were between $\bar{X} = 4.26$ and $\bar{X} = 1.87$. The item of "*There should be ready-made exams in the books to assess listening skills*" had highest mean score whereas the item of "*The results of assessment should be used to place students*" had the lowest mean score. Other items' mean scores were respectively; students of a foreign language should be assessed in the primary

school ($\bar{X} = 4.2$), there should be ready-made exams in the books to assess speaking skills ($\bar{X} = 4.12$), there should be ready-made exams in the books to assess listening skills ($\bar{X} = 4$), parents should always be informed with the results of students' assessment ($\bar{X} = 4.07$), I need to be more trained in assessing students in primary schools ($\bar{X} = 3.97$), students should be assessed to check students strength and weaknesses for lesson planning ($\bar{X} = 3.68$), students should be tested on sentence-level activities ($\bar{X} = 3.68$), students should be assessed by comments, not by numerical grades ($\bar{X} = 3.29$), it is difficult to assess speaking skills in the crowded classes ($\bar{X} = 3.09$) and students should be assessed to provide a record for success and failure ($\bar{X} = 2.92$).

Table 31

General Beliefs about Assessment of Students' Foreign Language (Turkey)

General Beliefs about Assessment of Students' Foreign Language (TURKEY)											
<u>Items</u>	\bar{X}	FA		A		PA		DA		DAA	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Students of a foreign language should be assessed in the primary school.	3.26	18	20.2	14	15.7	39	43.8	10	11.2	8	9
I need to be more trained in assessing students in primary schools.	3.48	14	15.7	35	39.3	22	24.7	16	18	2	2.2
It is difficult to assess speaking skills in the crowded classes.	4.68	65	73	21	23.6	2	2.2	1	1.1	-	-
There should be ready-made exams in the books to assess listening skills.	4.19	33	37.1	43	48.3	11	12.4	1	1.1	1	1.1
Students should be assessed to check students' strength and weaknesses for lesson planning	3.98	29	32.6	39	43.8	15	16.9	3	3.4	3	3.4
Students should be assessed by comments, not by numerical grades.	3.66	20	22.5	30	33.7	30	33.7	7	7.9	2	2.2
The results of assessment should be used to place students.	2.97	10	11.2	22	24.7	24	27	22	24.7	11	12.4
There should be ready-made exams in the books to assess speaking skills.	3.93	28	31.5	34	38.2	20	22.5	7	7.9	-	-
Students should be assessed to provide a record for success and failure.	3.24	13	14.6	26	29.2	28	31.5	14	15.7	8	9
Students should be tested on sentence-level activities.	3.15	8	9	28	31.5	31	34.8	14	15.7	8	9
Parents should always be informed with the results of students' assessment.	3.97	30	33.7	35	39.3	17	19.1	6	6.7	1	1.1

FA : I fully agree, A: I agree, PA: I partially agree, DA: I don't agree, DAA: I don't agree at all

Table 32

General Beliefs about Assessment of Students' Foreign Language (Italy)

General Beliefs about Assessment of Students' Foreign Language (ITALY)											
<u>Items</u>	\bar{X}	FA		A		PA		DA		DAA	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Students of a foreign language should be assessed in the primary school.	4.2	13	32.5	22	55	5	12.5	-	-	-	-
I need to be more trained in assessing students in primary schools.	3.72	12	30	14	35	9	22.5	1	2.5	4	10
It is difficult to assess speaking skills in the crowded classes.	3.7	13	32.5	13	32.5	7	17.5	3	7.5	4	10
There should be ready-made exams in the books to assess listening skills.	4.15	11	27.5	25	62.5	3	7.5	1	2.5	-	-
Students should be assessed to check students' strength and weaknesses for lesson planning	4.15	14	35	19	47.5	6	15	1	2.5	-	-
Students should be assessed by comments, not by numerical grades.	3.62	11	27.5	12	30	11	27.5	3	7.5	3	7.5
The results of assessment should be used to place students.	3.45	2	5	20	50	12	30	6	15	-	-
There should be ready-made exams in the books to assess speaking skills.	4.15	14	35	19	47.5	6	15	1	2.5	-	-
Students should be assessed to provide a record for success and failure.	3.67	6	15	19	47.5	11	27.5	4	10	-	-
Students should be tested on sentence-level activities.	3.7	2	5	26	65	10	25	2	5	-	-
Parents should always be informed with the results of students' assessment.	3.92	14	35	14	35	8	20	3	7.5	1	2.5

FA: I fully agree, A: I agree, PA: I partially agree, DA: I don't agree, DAA: I don't agree at all

Table 33

General Beliefs about Assessment of Students' Foreign Language (Finland)

General Beliefs about Assessment of Students' Foreign Language (FINLAND)											
<u>Items</u>	\bar{X}	FA		A		PA		DA		DAA	
		f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)	f(n)	p(%)
Students of a foreign language should be assessed in the primary school.	3.97	13	31.7	16	39	11	26.8	-	-	1	2.4
I need to be more trained in assessing students in primary schools.	3.09	6	14.6	11	26.8	7	17.1	15	36.6	2	4.9
It is difficult to assess speaking skills in the crowded classes.	4	17	41.5	11	26.8	9	22	4	9.8	-	-
There should be ready-made exams in the books to assess listening skills.	4.26	18	43.9	17	41.5	5	12.2	1	2.4	-	-
Students should be assessed to check students' strength and weaknesses for lesson planning	3.68	8	19.5	14	34.1	17	41.5	2	4.9	-	-
Students should be assessed by comments, not by numerical grades.	3.29	4	9.8	14	34.1	13	31.7	10	24.4	-	-
The results of assessment should be used to place students.	1.87	-	-	1	2.4	7	17.1	19	46.3	14	34.1
There should be ready-made exams in the books to assess speaking skills.	4.12	15	36.6	17	41.5	8	19.5	1	2.4	-	-
Students should be assessed to provide a record for success and failure.	2.92	3	7.3	10	24.4	13	31.7	11	26.8	4	9.8
Students should be tested on sentence-level activities.	3.68	4	9.8	23	56.1	11	26.8	3	7.3	-	-
Parents should always be informed with the results of students' assessment.	4.07	16	39	13	31.7	11	26.8	1	2.4	-	-

FA: I fully agree, A: I agree, PA: I partially agree, DA: I don't agree, DAA: I don't agree at all

In order to examine whether teachers' general beliefs about assessment of 4th grade students' foreign language changed according to their country, one way analysis of variance (ANOVA) was performed. ANOVA results are given in Table 34.

Table 34

ANOVA Results in Terms of Teachers' General Beliefs about Assessment of 4th Grade Students' Foreign Language

ANOVA Results in Terms of Teachers' General Beliefs about Assessment of 4th Grade Students' Foreign Language					
<u>Country</u>	<u>N</u>	<u>\bar{X}</u>	<u>SD</u>	<u>F test</u>	<u>Comparison</u>
Turkey	89	40.57	5.99	F (2, 167) = 3.962 , $p < 0.05$	Italy > Finland
Italy	40	42.45	5.62		
Finland	41	39.0	4.2		

As presented in Table 34, the results of ANOVA showed that teachers' general beliefs about assessment of 4th grade students' foreign language differed significantly in terms of country [F (2, 167) = 3.962 , $p < 0.05$]. Follow up analysis using post hoc comparison with Tukey test was conducted to observe pair-wise differences among the countries. The post hoc analysis showed that Italian teachers' general beliefs about assessment of 4th grade students' foreign language ($\bar{X} = 42.45$, SD = 5.62) was significantly higher than Finnish teachers' general beliefs ($\bar{X} = 39.0$, SD = 4.2).

4.2. Interview Results

Under this title, the results revealed from content analysis of interviews with teachers in three countries were presented across their countries. In the presentation of the results, the codes emerged from the content analysis were supported with quotations taken from individual interviews. In giving quotations, teachers' real names were not directly given. Instead, their names were coded as T_EA, I_CF and F_TW. The first letter refers to the

country (T-Turkey, I-Italy and F-Finland), the other letters refer to the initials of teachers' real names.

4.2.1. Exams used for young learners. The teachers were asked to indicate their thoughts regarding the exams they use for young learners and appropriateness of the types or exercises of the exams for the young learners. They were also asked to indicate whether the exams or exercises they used assess the right skills. Teachers in different countries reported various responses to these questions. Some believed that the exams and exercises were appropriate and assessed the right skills, but some others said that they were not enough.

4.2.1.1. Turkish teachers. Some of Turkish teachers thought that the exams they used were appropriate, but some others did not believe in their appropriateness for young learners.

Teacher T_EA explained how she assessed their students and the process of assessing the young learners as

“I have preferred performance-based assessment for the second and third grades to evaluate their listening and speaking skills. I believe that is the correct way of evaluation as the literature suggests, but I don't approve of the fourth grades' evaluation criteria as stated in the curriculum. As a teacher, I am expected to evaluate their reading and writing skills as stated in the curriculum and I feel obliged to do it.”

Teacher T_DA believed in the appropriateness of the exams they used for young learners and reported that

“They are appropriate for the students to some extent. Because of time limitation and other reasons, I usually omit listening and writing parts of the exams.”

One teacher found the exams appropriate for young learners, but not enough in terms of assessing communicative skills. The other teacher explained the examination process that he tended to prepare easy exams to let the students get higher scores which would result in higher motivation.

On the other hand, many of the teachers reported some weaknesses stemming from the exam itself and external effects. The weaknesses that make the exams inappropriate for the young learners are associated with time, the content of the exams (e.g. focusing more on vocab and reading, not speaking and listening; assessing content more), diversity of skills (e.g. having difficulties in including all types of skill into the exams) and length of the exams (e.g. long exams requiring more time to respond).

Teacher T_DK did not believe in the appropriateness of the exam to the young learners and explained the reason as

“...Because we use written exams that evaluate mostly vocabulary knowledge and reading comprehension. We can't evaluate speaking and listening skills.”

4.2.1.2. Italian Teachers. Italian teachers believed that the exams they used were appropriate for assessing young learners, but they emphasized on the need for target exercises and modification of the exams.

Teacher I_AM stated that they used the exams of Cambridge University for Young Learners (YLE, STARTERS, MOVERS, Level Pre A1, A1 CEFR) for a long time and found that type of examination very appropriate for the age group of 8 – 11 and useful in terms of students' intercultural, communicative and language skills since they reinforced pupils' self-confidence and capacity to orient.

4.2.1.3. Finnish teachers. Similar to Turkish and Italian teachers, some Finnish teachers found the exams appropriate, but some others did not find them appropriate for assessing young learners.

The teacher F_KP had some difficulties for the exams and said that

“I have some difficulties in evaluating speaking skills. Oral exams are difficult to organize as they take up so much time with 80 students”.

Teacher F_SMS reported that the exams assessed limited types of skills of young learners and said that

“Exams evaluate the language skills but not other skills which make most of the national curriculum aims for language teaching (language and culture awareness, learning to learn, general communication skills). Actual language skills only make a fifth of all goals.”

Teacher F_SL reported that the types or items of the exams were partly appropriate for the young learners and thus F_SL modified the ready-made exams for better versions.

Teacher F_TW found the Finnish textbook series for primary schools easy and handy and said that

“Finnish text book series for primary (elementary) education normally have test/exam sets for different skill levels (basic, intermediate and advanced) from which you can choose test tasks (items) and modify exams suitable for each group.”

4.2.2. The ways of assessing communicative skills. The teachers were asked to explain how they assessed communicative skills and whether the ways they used were effective.

4.2.2.1. Turkish teachers. Turkish teachers tried to improve their students' communicative skills with the help of classroom activities, formative assessment, rubrics, scales for assessment, improving imagination, asking students to interpret what they saw, listened or drew, doing pair-works, question-answers, listening exams and in-class participation.

Teacher T_ÖC expressed some ways to assess very limited part of communicative skills and did not find the exams effective for young learners. She explained this as

“... We have a listening exam, but actually we give the students Cambridge movers exams as listening exams. I don't think they are very effective because the students

who cannot understand what I am saying in class can get high scores in those exams. As for speaking, we don't officially assess them, but as a teacher I just observe the students and so I have an idea about their speaking skills.”

4.2.2.2. Italian Teachers. Italian teachers stated that they used various numbers of exercises and procedures to assess communicative skills of young learners. The ways they used to assess communicative skills were individual tests, group games, conversation, TPR approaches, specific and practice-based exercises, interviews with students, encouraging students to communicate with their peers in other countries, books, video conferences, creating an intriguing and stimulating learning.

Teacher I_CF made students talk more, often asked questions to get them practice the language to improve their communicative skills, but complained about the course hours in a week. I_CF thought that the course would be more effective if the course hours had been 3 hours in a week. Teacher I_SB assessed the students' communicative skills through using TPR and reported the way she performed as

“Total physical response. Children are involved, they repeat the words and sounds associating them with a movement or gesture.”

4.2.2.3. Finnish teachers. Finnish teachers reported several ways to assess communicative skills. The ways they used for assessing young learners' communicative skills were including listening parts in the exams, small exercises, nation-wide tests, listening comprehensive tests, reading / speaking / writing / pronunciation tests, verbal tests and teacher observations.

The teacher F_KP indicated that assessment of speaking skills was much more difficult compared with other skills and thus said that

“I have some difficulties in evaluating speaking skills. Oral exams are difficult to organize as they take up so much time with 80 students.”

Teacher F_TW reported to use teacher observations for assessing communicative skills and use exams for communicative skills, but thought that

“Those are a bit demanding to administer in the normal course of lessons (one or a few to be tested, the rest left alone)”

4.2.3. Using exams that assess communicative skills. The teachers were asked to report whether they would use exams for assessing communicative skills more if they found them ready-made in the course book or on an online platform.

4.2.3.1. Turkish teachers. Most of Turkish teachers remarked that they would use ready-to-use exams taken from the coursebooks or on an online platform. One teacher reported that she was loyal with the coursebook, the other one preferred to use ready-made exams since it saved time.

Teacher T_DH tended to use any types of exams depending on their quality and appropriateness for the students she taught. Teacher T_AE believed in the assistance of coursebooks as well as online platforms to assess students and said that

“Coursebooks provide great assistance for us because students can see the questions and activities there. It helps them understand the sentences before I ask them questions. Online platforms are also beneficial because I can use different types of exercises there.”

Teacher T_EÇ preferred to use online exams depending on the facilities of the schools and also to use coursebooks, and said that

“I would use exams on an online platform but it actually depends on the technological facilities of the school. If classes have smart boards and books have interactive white board software, we can mostly use coursebooks. Because these books integrate all skills. Also, they have lots of visual contents. This is a very important point. Because young learners can learn by seeing and acting out what they see.”

4.2.3.2. Italian Teachers. Many teachers expressed that they could use exams more if they found them ready. Even though they tended to use ready-made exams that took place in the course books and on online platforms, one teacher indicated to use these exams if they were in line with the learning objectives while the other teacher stated to use these exams by adjusting them according to the students' level of understanding.

Teacher I_SC used ready-made exams but said that “I like to choose them and eventually adjust for my students”.

Similarly, Teacher I_SB preferred to use ready-made exams but assess communicative skills in action and said that

“Well, personally yes, but seeing those skills in action is more effective than books. For example, I had The Bell Beyond teachers acting in my school and I learnt a lot from their TPR method.”

Teacher I_AEH stated that he would use ready-made exams in the course books and on online platforms since he felt assisted, but he would add extra materials to match the content / skills taught in the classroom. Teacher I_ABC reported to use ready exams as much as they fit the learning objectives.

4.2.3.3. Finnish teachers. Finnish teachers used course book materials, computer based materials and also their own materials for the assessment of communicative skills of young learners.

Teacher F_TW reported to use computer simulation for communicative purposes, but asserted the difficulties to access this simulation as

“I have experimented on computer simulations for (also) communicative purposes and they were great (e.g. Wewer 2014)! However, they are expensive and not within everyone's reach. So, online platforms would be very appealing if they were part of a coursebook provision.”

4.2.4. Using language portfolio as an assessment tool. Teachers were asked to indicate their thoughts of using language portfolio as an assessment tool in their FL classrooms and its usefulness.

4.2.4.1. Turkish teachers. Almost all teachers mentioned the usefulness of language portfolio for the learners, but some teachers thought that despite its benefits, it was not appropriate especially for the young learners. They found the language portfolio useful because it

- helped to evaluate students' progress easily,
- helped students be aware of their own progress,
- supported self-esteem,
- developed sense of responsibility,
- provided permanent learning,
- helped students to try to be better than their peers and themselves and
- was appropriate in case of limited class time.

One teacher said that it was appropriate for crowded classrooms, however the other one believed in its usefulness in case of having fewer students in the class.

One of Turkish teachers who believed in the necessity of language portfolio for young learners pointed out that

“It’s necessary because the children cannot reflect their communicative skills with paper and pencil tests. As teachers, we have a chance to observe how they perform during the whole term. Therefore, we should use portfolios. I use digital portfolio in classdojo app, so that I can easily evaluate my students' development.” (Teacher T_DA).

Teacher T_DK believed in the usefulness and necessity of the language portfolio because of its contribution to the assessment of the process, but he did not find it appropriate

for young learners because of the difficulty to use it with young learners. The other teacher T_EC did not believe in the usefulness of language portfolio due to the fact that

“...Parents come to school to watch their children. Before the presentation, student’s file is prepared which includes chosen activities and they are repeated many times. Students are stressful. Because it is like a show and it must be perfect, it is not student’s real performance. Schools consider it like an advertisement”.

4.2.4.2. Italian Teachers. Some Italian teachers found language portfolio useful and beneficial, but some others found it not necessary especially for young learners. The ones who found it useful thought that it had positive effects, it influenced students’ progress in a positive way, it helped to learn more words and find everything faster and it provided authentic practice, autonomy, motivation for students.

Teacher I_AM found using language portfolio useful because

“Portfolio is an authentic assessment practice serving learners and their families. It is based on learner’s autonomy, motivation. It has to be practiced with continuity and coherence throughout the study cycle, even in one of its components: Autobiography, Dossier, PassPort.”

4.2.4.3. Finnish teachers. Using language portfolio was found to be useful by many, but some others notified that it may not be useful. One teacher said that it was useful, but not appropriate for young learners due to the fact that it focused more on oral skills. The other teachers found using language portfolio not useful since it forced young learners. On the other hand, one teacher found using language portfolio useful because it made the language learning tangible.

4.2.5. Thoughts on self and peer-evaluation. Teachers were asked to indicate their thoughts about using self or peer-evaluation as an assessment and their positive and/or negative sides.

4.2.5.1. Turkish teachers. Many of the teachers believed in benefits of self or peer-evaluation, but some teachers reported that it was not useful and had some negative effects at this age. The ones who believed in its appropriateness claimed that it

- helped students to observe how well they and their peers learned the subject,
- helped to see mistakes
- encouraged the students
- prevented false learning
- developed respect to others
- helped to teach self-consciousness
- helped to teach responsibilities and cooperation
- helped students be autonomous

On the other hand, some teachers stated the weaknesses or inappropriateness of the peer assessment due to a cultural shortcoming, being cruel to others, tending to accuse each other in case of failure, being hard on each other, being rude while criticizing, being not objective and over judge themselves.

Teacher T_{SG} did not see these forms of assessment appropriate and explained this as “I think that in our country, self or peer-evaluation as an assessment cannot be objective. This is a cultural shortcoming”

Similarly, Teacher T_{EÇ} did not think that self-evaluation was not appropriate for the young learners because of their lack of ability to criticize and give feedback by themselves. On the other hand, she found peer-assessment more appropriate for giving feedback even though it still had some negative effects since the students might show cruelty to the others’ thoughts and actions and liked to talk about their weaknesses.

4.2.5.2. Italian Teachers. Italian teachers reflected that self or peer-assessment was useful and had positive effects because especially self-evaluation was stimulating, helped the students to think about what they had learned and enhanced students' performance.

Teacher I_AM supported the effectiveness of self or peer-assessment by saying that "Self and peer evaluation are other very effective components of enhancing learners' performances. It should be a circular, natural path to be evaluated, to be assessed, to evaluate, to assess. It requires a sense of responsibility and balance. Students can find it very funny and stimulating but the teachers must always supervise, regulate, adjust the process, if needed".

4.2.5.3. Finnish teachers. Finnish teachers thought affirmative about using peer - or self – evaluation as an assessment procedure since they eased up the work load and motivated the students, but they still needed to be used carefully so as to be functional and beneficial. In terms of making peer assessment more useful and practical, teacher F_TW believed that

"Peer evaluation should be well guided (criteria, means of giving feedback) before actual use. Learners seem to value teachers' direct feedback more than that of their peers."

4.2.6. Difficulties in assessing young learners. Teachers were asked to specify the difficulties they faced in assessing young learners.

4.2.6.1. Turkish teachers. Turkish teachers reported many difficulties they faced during the assessment of young learners. The difficulties they encountered were being shy to speak, lack of vocabulary, time limitation, classroom size (e.g. crowded classroom), not having technological devices, inadequacy of coursebooks in terms of communicative skills, having prejudices against foreign language, being reluctant and shy, having too many questions in a simple exam / task and challenge of assessing reluctant students.

4.2.6.2. Italian Teachers. Italian teachers also presented some difficulties they faced in assessing young learners. These difficulties were limited class hours, students' capacity to talk, feeling embarrassed, having foreign kids in the class, improperly designed books, grammar, giving marks, difficulty in preparing different assessment tools / models for each student (individual differences) and parent perceptions. One teacher stated her difficulty as "assessing speaking skills because other students can't wait for a long time." (Teacher I_SC)

4.2.6.3. Finnish teachers. Finnish teachers mentioned several difficulties they encountered while assessing young learners such as lack of time, classroom size, reaching all students in the class (face to face interaction), too much emphasis on one skill (e.g. writing), preparing the materials and exams without assistance.

To teacher F_SMS, the difficulty came from the curriculum itself and its interpretation, and she reported that

"The national curriculum is very difficult to interpret, it's vague and makes a huge uncertainty factor. Am I doing/assessing the right thing in the right way? You can never be sure since it's a matter of your local interpretation of the national guidelines."

4.2.7. Solutions for the difficulties in assessing young learners. The teachers were asked to explain how they solved the problems they encountered during the assessment of young learners.

4.2.7.1. Turkish teachers. The solutions or the ways Turkish teachers proposed were encouraging the students to speak more, giving extra importance to learning new words, evaluating students in a small group in a short period, decreasing the number of the students in the class, using an efficient coursebook, training teachers, using technology and internet, doing the evaluation in groups, using enough materials and technological devices, and not forcing students to have exams.

4.2.7.2. Italian Teachers. Italian teachers proposed several solutions for dealing with the difficulties they encountered. These solutions were related to speaking English more, having an assistant, increasing the number of the class hours in a week, creating a peaceful and tolerant environment, sharing with and co-working within a team, increase the quality of the book, planning game and activities, planning the assessment procedure carefully, being patient and being clear on the expectations from the students.

Teacher I_CF stated that the capacity to talk and feeling embarrassed are the difficulties faced in assessing young learners, and thus tried to create a learning environment as peaceful and tolerant to mistakes as possible to solve problems.

4.2.7.3. Finnish teachers. Finnish teachers presented some solutions such as using more verbal exams, making multi-faced assessment, drawing from multiple sources and focusing on different group of students in each lesson time.

4.2.8. Types of assessment in case of proper conditions. The teachers were asked to report the types of assessment they would use if they had proper conditions. The types of assessment that teachers indicated to use in proper conditions were given in Table 35 comparatively.

Table 35

Types of assessment teachers would use if they had proper conditions

<u>Turkish teachers</u>	<u>Italian Teachers</u>	<u>Finnish teachers</u>
<ul style="list-style-type: none"> • Interactive and computer-based evaluation • exams including all four of the language skills • listening to a text or dialogue and answer the 	<ul style="list-style-type: none"> • small and frequent test, • assessing the way students speak to each other, • doing assessment in 	<ul style="list-style-type: none"> • mixed evaluation competencies (listening, writing, reading, grammar, function, digital competence etc.),

- | | | |
|------------------------------|--------------------------|---------------------------|
| questions accordingly | action such as | • combination of oral and |
| • watching cartoons or | speaking, drama, pair- | written exams and |
| videos appropriate for | work and during | exercises, |
| students and letting the | games, | • more oral exams |
| students talk about them | • all skills together in | • computer simulations. |
| and discuss. | an assessment, | |
| • making education | • listening tests, | |
| enjoyable | • checking their | |
| • exams including not just | abilities, | |
| grammar but listening | • form filling, | |
| and speaking skills | • self- , peer- and | |
| • doing activities in groups | group- assessment, | |
| • portfolio | • reading | |
| • assessing process and | comprehension test, | |
| progress | • real tasks | |
| • observation | • interaction with real | |
| | interlocutors. | |

4.2.9. The ways to improve assessment skills. Teachers were asked to indicate the ways to improve their assessing skills.

4.2.9.1. Turkish teachers. The ways that Turkish teachers proposed to improve assessing skills were using teacher forum, attending seminars or webinars, observing the students, reading more, searching and using different types of exercises in the exams, in-service trainings, more experience through discussing with colleagues and sharing experiences among teachers.

4.2.9.2. Italian Teachers. Italian teachers suggested several ways to improve their own assessment skills of young learners. The ways to improve assessment skills were conferences, online courses, observing students, courses, exchanging ideas with other teachers, buying additional materials for enhancing the lesson, planning in advance, receiving feedback from experienced colleagues, experiencing abroad, participating in continuous trainings, Erasmus plus projects and practicing reflective professional self-assessment.

4.2.9.3. Finnish teachers. Finnish teachers offered several ways to improve their assessment skills. These ways were using web tools, working hard, following the guideline in the curriculum, receiving more training, comparing their work with others, discussing educational issues, adopting new skills, dedicating more time for planning the assessment, experimenting different scheme (e.g. group assessment) and being aware of the criteria and objectives in the curriculum.

4.2.10. General suggestions for assessment of young learners. At the end of the interview, teachers were asked to share their comments or suggestions to assess young learners. Turkish teachers provided much more comments on the assessment of young learners than the teachers in Italy and Finland.

The suggestions proposed by Turkish teachers were as follow

“Teachers of young learners have to know very well and take into account the qualities of young learners while assessing them. The course books should include more spoken evaluation forms and revision questions for oral assessment.” (Teacher T_DA)

“In my country, the Ministry of Education should provide more training on assessment methods. While teaching and learning methods change, evaluation methods differ as well. I think, we need continuous improvement.” (Teacher T_EA)

“I wish they didn’t get grades according to their exam results. They are too young to get nervous about grades.” (Teacher T_ÖC)

The suggestions proposed by Italian teachers were as follow

“It would be amazing to be 2 teachers to organize activities in the best possible way” (Teacher I_AEH)

“Formative assessment more than summative” (Teacher I_ABC)

“More hours in Italy” (Teacher I_SC)

Only one teacher from Finland provided comments on the assessment and added that

“Assessing children should not play a big role in teaching. It's more important to use time for learning the language and especially find useful ways to use it.” (Teacher

F_SL)

Chapter 5

Discussions and Conclusion

This study was carried out to investigate foreign language teachers' practices and beliefs about the ways and types of assessment tools, assessment activities and exercises used for young learners' foreign languages in Turkey, Italy and Finland. Both qualitative and quantitative data were collected through the use of questionnaire and individual interviews, so as to address to the following research questions;

1. How is the assessment being implemented by EFL teachers at 4th grades in primary schools in Turkey, Finland and Italy?
2. What are the assessment activities used by EFL teachers in Turkey, Finland and Italy?
3. What are the content and related skills in assessment in Turkey, Finland and Italy?
4. What are the teachers' beliefs about assessment of 4th grade students in Turkey, Finland and Italy?
5. What are the problems in the assessment practices in Turkey, Finland and Italy?
6. What are teachers' deficiencies in the assessment of young learners in Turkey, Finland and Italy?

In this chapter, the major findings revealed in the study were reported and discussed across national curricula and previous studies. At the end of the chapter, the limitation of the study and implications for educational practice and further research were presented.

5.1. Discussion

Assessment has recently received greater attention and had critical importance in ELT (Davison & Cummins, 2007) due to its contribution to language policy and its design. As a part of language teaching process, teachers use assessment in several ways and types, and also

for different purposes. This study showed teachers' various assessment practices and beliefs in three countries in detail.

Teachers' reasons to use classroom assessment for young learners varied across countries. Even though teachers stated different reasons to carry out classroom assessment, the most frequent reasons indicated by teachers were different. Turkish teachers tended more to use classroom assessment for providing a record to students, Italian teachers for checking strengths and weaknesses, and Finnish teachers for providing information to parents and students. In addition to these frequently reported reasons, teachers also reflected other reasons. Along with all these reasons, Turkish teachers used classroom assessment mostly for summative and diagnostic purposes and motivation, Italian teachers mostly for diagnostic purposes and feedback and Finnish teachers for diagnostic purposes, feedback and motivation. The distinction among the teachers referred that Turkish teachers inclined to use assessment generally for summative purposes whereas Italian and Finnish teachers used assessment for formative purposes. Doğandere (2006) asserts that the basic aim of assessment in language learning is to provide adequate feedback in order to help learners monitor their strengths and weaknesses in FL. Similarly, Harmer (2001) believes in the importance of teacher feedback due to its positive effect on students' language use in the future.

The first research question addressed to the ways / types of assessment being implemented in EFL teachers at 4th grades in three countries. Turkish and Finnish teachers mostly used their own – made paper and pencil test whereas Italian teachers mostly used oral interview to assess 4th grade students' foreign language. On the other hand, peer assessment in Turkish EFL classes, language portfolio in Italian EFL classes and standardized paper and pencil tests from national or local EFL education authority in Finnish EFL classes were the least used assessment types. Examining their beliefs about types of assessment to be used at the 4th grades, Turkish teachers agreed to use performance-based test and oral interview,

Italian teachers agreed to use oral interview and Finnish teachers agreed to use their own-made paper pencil tests, oral interviews and self-assessment. In Turkish FL classes, a mismatch was observed between what teachers implemented (e.g. their own-made paper and pencil test) and what they tended to implement (highest beliefs scores on performance-based test and oral interview). In Finnish FL classes, the assessment types that teachers mostly implemented (mostly their own-made paper and pencil test) was observed to be partially in line with what they tended to implement (highest beliefs scores on their own-made paper and pencil test performance-based test, self-assessment and oral interview). A complete match was observed in Italian FL classes in terms of what assessment ways was being implemented and what assessment ways to be implemented. As for the ways for assessing communicative skills of young learners, teachers used a variety of tools and techniques. Rather than standardized tests, they used assessment tools for formative purposes and encouraging students to involve in communication process through talking, listening, pair-works etc.

Even though paper-pencil tests are used widely in FL classes, these tests should be carefully designed and used since they can affect young learners negatively (Pinter, 2009). As claimed by Conner (2008), poorly designed and applied paper-pencil tests can lead to learners losing interest and motivation in language learning, can be stressful and can influence the effective use of learning time. Furthermore, paper-pencil tests can affect young learners' self – image as well as increasing the level of anxiety (Smith, 1996). On the other hand, performance tasks that teachers reported to be used for language classes provide benefits in terms of vocabulary improvement, grammar reinforcement, increased motivation and confidence (Yıldırım & Orsdemir, 2013). Similarly alternative assessment techniques also enhance learner motivation and active participation (Kohonen, 1997).

The results of previous research on the same topic were observed to be somehow parallel with the current study. In a study done in Finnish schools, Wever (2014) reported that

the most commonly used assessment types in language classes were teacher observation, dialogic interaction and bilingual tests whereas the least used assessment types were peer-assessment, simulations and portfolios. Önalın and Karagül (2018) found that Turkish EFL teachers tended to use assessment for formative purposes and self-assessment was an important procedure in EFL classes. In another study implemented with Turkish EFL teachers, Kirkgoz, Babanođlu and Ađçam (2017) reported that teachers preferred to use performance-based and communication-based assessment types for assessing students' success, but most frequently used assessment tools were paper-pencil tests, in-class observations and performance tasks in language classes. The differences among the teachers in terms of the most and the least frequent usage of assessment types could be due to the teachers' curriculum orientation, their approach toward teaching and assessment, being more stick to national curriculum and the books.

Even though teachers found language portfolio to be useful in terms of its contribution to assessment of students' progress (e.g. language, personal attributes, motivation and so on), they believed in its inappropriateness for young learners. Zorba and Tosun (2011) conducted a study with kindergarten students and found that language portfolio and additional instructional materials helped to increase young learners' motivation, prolonged attention span and involvement in the course. Gonzales (2009) emphasized the importance of using portfolio for young learners as well. Teachers preferred to use self-assessment rather than peer-assessment because of its more positive sides. They believed that self-assessment helped learners monitor their success and progress. In the study done with EFL teachers, Tatlı (2014) found that teachers assigned performance and project tasks to young learners rather than self-assessment and portfolios since they believed in students' lack of skills, objectivity and autonomy to reflect and assess their own learning. In this regard, McKay (2006) claimed that young learners should be supported when they were involved in self-assessment.

The second research question addressed to the assessment activities / exercises used at 4th grade FL classes by EFL teachers. Frequently used assessment exercises in Turkish FL classes were matching, filling gaps, comprehension questions about a short text, multiple choice and grammatical activities; in Italian FL classes were comprehension questions about a short text, filling gaps, matching, listening to audio materials, multiple choice tests, writing words, short texts, reading words and sentences, grammatical activities and oral description of a picture; and in Finnish classes, writing words, short texts, filling gaps, grammatical activities, matching, comprehension questions about a short text and students' creation of their own sentences / texts. The activities of dictation was the least used assessment activities in Turkish and Italian FL classes and the repetition of vocabulary or sentences was the least used in Finnish FL classes. These results indicate that Turkish EFL teachers preferred to use traditional exercises more in their classrooms. On the other hand, Italian and Finnish teachers used a variety of assessment activities in a more balanced way. Rather than using one or two activities dominantly, using various assessment activities in balance could also enable teachers to assess students from different edges at different learning levels (cognitive, affective etc.). Since young learners (especially children) are mentally active and want to understand the meaning and rationale of the activities designed to them (Cameron, 2001), more meaning and explanation related to the activities and more relaxed environment should be provided (Tatlı, 2014). In the study undertaken with teachers teaching English to Turkish young learners, Ayas (2014) found that the most common preferred assessment practices in FL classes were teacher-made written test, oral questions – answers and planned observation whereas portfolio, self / peer – assessment, student written work, standardized tests, essays, unplanned observation and conferencing were relatively less used. Kırkgöz and Ağçam (2012) conducted a study for examining the written assessment practices of English teachers in primary schools in Turkey and found that constructed response items were more

predominant in written assessment practices at 4th grade. Öz (2014) found that the most frequent assessment methods that Turkish EFL teachers used were fill-in-the-blank, multiple choice, true-false, matching and short-answer questions whereas the least frequently used assessment methods were rubrics, self- and peer-assessment, observation, drama and structured grid.

The third research question addressed the content and related skills assessed at 4th grade FL in assessment. Compared with other language content / skills, lexis (vocabulary) was reported to be frequently assessed content / skills at 4th grade FL classes in three countries. Pronunciation in Turkish and Finnish FL classes, and spelling in Italian FL classes were relatively the least assessed content / skill. In addition to lexis (vocabulary), reading skills, grammar and writing skills in Turkish classes; speaking skills, listening, reading, writing, grammar and pronunciation in Italian classes; grammar, listening skills, reading skills, writing skills and spelling in Finnish classes were other frequently used content and skills. The skills focused in English language curriculum at 4th grade in Turkey are “listening, speaking, very limited reading and writing (p.10)” (MEB, 2018). Furthermore, language curricula and books of three countries are designed in line with common European Framework of Reference for Languages: Learning, Teaching and Assessment (CEFR). CEFR developed by the council of Europe in 2001 for providing common language framework across Europe (Council of Europe, 2001) have been considered in Turkish (MEB, 2018), Italian (Faez, 2011) and Finnish language curriculum (Helminen, 2013). However, Turkish FL curriculum aims 4th graders to have language skills at A1 level. On the other hand, A2 syllabus for English is applied to 4th graders in Finland. The diversity in content and skills assessed at 4th grade FL classes is mainly due to the course content and curriculum itself. Different countries’ FL curriculum put emphasis on the same language content and skills, but at different degree based on national language policy. Of course this difference is reflected in

course implementation and thus assessment of students' language learning. Genesee (2001) and McNamara (2000) claimed that to perform authentic tasks for assessment, four skills of language learning were needed to be used and reflected in assessment procedures.

The fourth research question addressed to the teachers' beliefs about assessment of 4th grade students. Turkish teachers agreed that performance-based tests and oral interviews as assessment tools; speaking skills, listening skills, lexis (vocabulary), reading skills and pronunciation as content / skills in assessment; and audio materials, matching, oral description of a picture, role playing, making a dialogue, comprehension questions about a short test, filling gaps as the kinds of assessment were needed to be used at 4th grade language classes. Italian teachers agreed that oral interview as an assessment tool; speaking skills, listening skills and reading skills as content / skill – focus in assessment; and listening to audio material, matching, oral description of a picture, role playing, making a dialogue, comprehension questions about a short test and filling gaps were needed to be used at 4th grade language classes. Finnish teachers agreed that self-assessment, teacher own-made paper and pencil test and oral interview as assessment tools; lexis (vocabulary), listening skills, speaking skills, reading skills and writing skills as content / skill – focus in assessment; and audio materials, students' creation of their own sentences /texts, comprehension questions about a short test, matching and writing words, short texts, filling gaps and grammatical activities were needed to be used at 4th grade language classes. Turkish, Italian and Finnish teachers' beliefs about assessment revealed their tendency to use more student-centered and formative-based assessment procedures. Writing skills and associated assessment tools and exercises were observed to be given less attention by many of the teachers in all three countries. The reason for giving less or little attention to writing skills that writing skill is not the main focus of 4th grade language learning. Speaking and listening skills are more emphasized and the main focus in the language curriculum for 4th grade level. That is why

teachers showed higher beliefs about the tools, content – skills and communicative exercises associated with listening and speaking skills. However, Finnish teachers, compared with Italian and Turkish, put much more emphasis on writing skills along with associated tools, content and exercises. Teachers' beliefs and tendencies regarding the ways of assessment tools, the types of content-skills and exercises are in line with the language proficiency to be gathered in A1 and / or A2 level according to CEFR. Teachers' general beliefs about assessment of young learners' foreign language revealed that the item "the results of assessment should be used to place students" had the lowest belief scores for most of the teachers.

Teachers' beliefs about the assessment affect their assessment practices in FL classes. In the present study, there seems to be a close match between teachers' beliefs about the ways of assessment, content-skills and exercises, and their practices to some degree, but a mismatch was observed in some items. This is more observable in Turkish sample. It is proved in the current literature that teachers' perceptions and conceptions of the assessment influence the ways they use the assessment and their assessment practices (Davison & Leung, 2009). Similarly, Brown (2008) asserted that teachers' conceptual beliefs affected their assessment practices.

The fifth research question addressed the problems that teachers encountered in the assessment system. Teachers' difficulties in assessing young learners in language classes varied in different countries. This variation among the teachers stemmed from the differences in students' background, teachers' equipment, classroom size, textbooks and educational policy and curriculum. Teachers in different countries presented their problems that they encountered during the assessment of young language learners. The first problem was associated with the exams used for assessing young learners' language learning. The teacher who believed that the exam was not appropriate for young learners indicated that the exams

had limitations in terms of the content and skills assessed, the time and the length. The difficulties that Turkish teachers faced were associated with the students' lack of motivation and vocabulary knowledge, class hours, crowded classes, not well designed classes and coursebooks. Similar, but additional difficulties were reported by Italian and Finnish teachers. The difficulties that Italian teachers have were students' aptitude to talk, their shyness, having foreign students in the classroom, having difficulty in preparing various assessment tools and parents' perceptions. In addition to the lack of time, the class size, the difficulty in preparing assessment tools; Finnish teachers identified the difficulty of reaching all students. Britton (2015) asserts that implementation of assessment of learning in EFL is closely related to the context in which it is done and also connected with timing, learner characteristics (ages, stages of cognitive development) and curriculum.

In order to deal with the difficulties, Turkish teachers proposed to encourage students to speak more, to assess students in a small group, to decrease number of the students in foreign classes, and to utilize additional materials and technology. For dealing with the difficulties, Italian teachers focused more on the number of the students, additional assistance in a class, encouraging classroom environment, planning of assessment and teacher role. On the other hand, in order to solve the problems in language classes, Finnish teachers put much more emphasis on quality and quantity of the assessment procedures.

The sixth research question addressed teachers' deficiencies in the assessment. The deficiencies that teachers have in the assessment had a match in three countries. Most of the teachers indicated that they had training to prepare and perform assessment of students at primary schools only at university. In all countries, teachers' lack of assessment skills was a common deficiency which was needed to be improved. In order to improve their assessment skills, Turkish teachers proposed to use teacher forums, attend seminars or webinars, observe the students, read more, search and use different types of exercises in the exams, join in-

service trainings, have more experience through discussing with colleagues and share experiences among teachers. To improve their own assessment skills of young learners Italian teachers suggested several ways such as conferences, online courses, observing students, exchanging ideas with other teachers, buying additional materials for enhancing the lessons, planning in advance, receiving feedback from experienced colleagues, experiencing abroad, participating in continuous trainings, Erasmus plus projects and practicing reflective professional self-assessment. Finnish teachers offered various ways to improve their assessment skills. These ways were using web tools, working hard, following the guideline in the curriculum, receiving more training, comparing their work with others, discussing educational issues, adopting new skills, dedicating more time for planning the assessment, experimenting different scheme (e.g. group assessment) and being aware of the criteria and objectives in the curriculum.

5.2. Summary

The multiple analyses of the both quantitative and qualitative data revealed significant outcomes with regard to foreign language teachers' beliefs about assessment of young language learners and their practices of assessment in language classes in Turkey, Italy and Finland. Following conclusions have been reached throughout the study.

1. The reasons to carry out the assessment for young language learners differed in three countries. Turkish teachers carried out assessment mostly for providing a record of students' achievement for success and failure, Italian teachers mostly to check students' strengths and weaknesses for lesson plans, and Finnish teachers mostly to provide information to parents and students.
2. Turkish and Finnish teachers stated that foreign language teachers are responsible for preparing the exams whereas a group of language teachers prepare the exams in Italy.

3. Turkish and Italian teachers mostly used numerical grades whereas Finnish teachers mostly used grade and comments together as a way of reporting exam results.

4. Most of the teachers in all countries received additional training for assessment of young learners only at university.

5. For the types of assessment used in FL classroom, “my own – made paper pencil test” was the most frequently used assessment type by Turkish and Finnish teachers while “oral interview” was the most frequently used assessment type by Italian teachers. Italian teachers’ frequency of using various assessment types significantly differed from Turkish and Finnish teachers’.

6. For the content / skills – focus in assessment of students, lexis (vocabulary) was the most frequently focused content by the teachers in all of three countries. In addition to lexis (vocabulary), reading skills in Turkish classes, speaking skills in Italian classes and grammar in Finnish classes were relatively more focused compared with other types of skills and content.

7. The variety of content / skills that Italian and Finnish teachers assessed at 4th grade was significantly higher than Turkish teachers.

8. Regarding the kinds of exercises used in FL classroom, most frequently used exercise was “matching” in Turkish FL classes, “comprehension questions about a short text” in Italian FL classes and “writing words, short texts” in Finnish classes.

9. The kinds of exercises used by Italian teachers in FL classes was significantly higher than that of Turkish and Finnish Teachers.

10. The highest belief score about the types of assessment to be used for 4th grade students for Turkish teachers was related with the item of “performance-based test”, for Italian teachers was related with the item of “oral interview”, and for Finnish

teachers was related with the item of “my own-made paper and pencil test” and “self-assessment”.

11. The highest belief score about the content / skills needed to be assessed for 4th grade students for Turkish teachers was related with the item of “speaking skills”, for Italian teachers was related with the item of “speaking and listening skills” and for Finnish teachers was related with the item of “lexis (vocabulary)”.

12. The highest belief score about the kinds of exercises needed to be used for 4th grade students for Turkish, Italian and Finnish teachers was the same and related with the item of “listening to audio materials”.

13. Teachers’ beliefs about the types of assessment to be used, about the content and skills needed to be assessed and about the kinds of exercises needed to be used for 4th grade students did not differ significantly across the countries.

14. The highest belief score about assessment of young learners’ foreign language for Turkish teachers was related with the item of “It is difficult to assess speaking skills in the crowded classes”, for Italian teachers was related with the items of “Students of a foreign language should be assessed in the primary school” and for Finnish teachers was related with the item of “There should be ready-made exams in the books to assess listening skills”. Italian teachers’ general beliefs about assessment of young learners’ foreign language were significantly higher than Finnish teachers.

15. As indicated by the teachers in all three countries, the exams used for assessing young learners’ FL were seen to have some weaknesses in terms of skills and content assessed and the length, and thus they were suggested to be revised / modified.

16. The ways to assess communicative skills of young learners varied across countries. Teachers in all three countries focused on a variety of assessment practices for communicative skills for both formative and summative purposes.

17. Teachers in three countries tended to use ready-made exams drawn not only from text books but also from online platforms.

18. Considering the teachers in all three countries, even though some teachers believed in the usefulness of language portfolio in terms of authenticity, motivation, responsibility, evaluating students' progress and so on, some others were against to it and found it not useful due to its inappropriateness in crowded classes and its difficulty to use for young learners.

19. Self-assessment was seen to be more beneficial for young learners according to most of the teachers in all three countries, when compared with peer-assessment. Self-assessment was considered to help students monitor their own progress and development. On the other hand, using peer-assessment in language classes was not supported due to the fact that young learners might tend to accuse each other, be cruel and rude to others, not to be objective to peers etc.

20. Despite its variety from classroom to classroom in different countries, the difficulties teachers' face in assessing young learners in all three countries were mainly about learning environment (e.g. crowded classroom), instructional technology and materials (e.g. inadequacy of book and technological devices in the class), students (e.g. being reluctant to speak, shyness, prejudice to foreign language), parent perception, class hours and assessment tools.

21. Teachers in all three countries suggested using multiple assessment techniques and procedures including various content and skills to mainly monitor young learners' progress and development in target language.

5.3. Limitations of the Study

This study was carried out with only language teachers in Turkey and two European countries, Finland and Italy. Face-to-face interaction with Turkish teachers was seen in the administration of the questionnaires and interviews, but the same procedures were not implemented with Italian and Finnish teachers. They were contacted through internet (e.g. teacher platform) and were required to give their responses in that way. This could provide comfort to the participants, but their responses to interview questions especially were not that much detailed. Some teachers did not understand some of the questions. Communication through internet especially for interview limited the researcher to deepen participants' feelings and thinking since no additional and / or prompt questions were asked to them.

Teachers' practices and beliefs regarding the assessment practices in EFL classes were only based on self-report data which was not assured with real class observations. Thus, teachers' self-report might not totally reflect what was being assessed and might somehow reflect teachers' goodwill.

EFL teachers from only three countries were selected for addressing the research questions. This helped to understand teachers' practices and beliefs of these countries. The effect of school conditions, demographics and curriculum orientations of teachers on their practices and beliefs were not compared in this study due to limited data.

5.4. Implications and Suggestions for Further Studies

This study has revealed some critical implications for classroom practices and further research in terms of assessment of young learners.

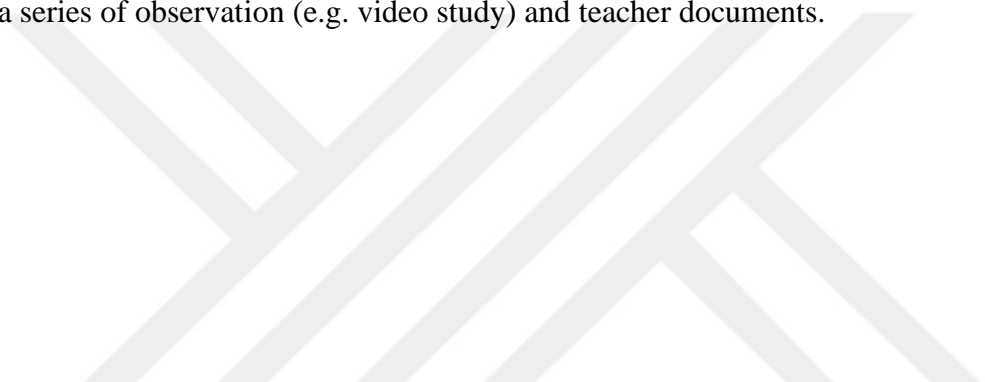
Even though teachers indicated to receive trainings on assessment of young learner at university years, they should be regularly taken to in-service trainings to refresh their knowledge and to make them aware of new trends, preparation of assessment tools and exercises. Furthermore, webinars, teacher conferences and seminars should also be organized

by local and national educational authorities to provide teachers to involve in discussions and experience-shared platforms. Also, teachers should be encouraged to join teacher platforms in the internet to share their experiences with colleagues in other schools and even countries to broaden their perspectives and learn the best practices. Besides these, teachers should make use of a variety of assessment procedures to assess student overall performance in learning foreign language. This could help teachers monitor and assess young learners' real performance since using one assessment procedure may be limited to assess learners' attainment in language.

Teachers stated several difficulties in relation to the assessment of young language learners. The common problems they faced were related to students' shyness and reluctance in learning English. In order to motivate students and cope with students' shyness, more motivating and intriguing classroom environment should be designed by encouraging students to involve actively in classroom practices, assuring them not to be graded, using pair-works, games, interactive activities, small group activities, videos and helping them speak more without fear of group evaluation. Furthermore, class size (number of the students) should also be decreased so that teachers can interact with students one-by-one and have more time to encourage them to speak. Course hours for language classes for 4th grade are very limited. Considering the whole class hours in a week, the time dedicated to language classes may be increased or if this is not possible, the course content especially for young learners should be reconsidered. The course books should be re-designed in terms of course content, skills to be emphasized, A1 level of CEFR and attractiveness.

For further research, more teachers from more different countries should be invited to a similar study to portray teachers' practices and beliefs in Turkey and different European Countries. Even though European countries follow CEFR standards for language learning, they still abide by their countries' educational policy. Comparing different countries in terms

of teachers' assessment practices and beliefs could also help to understand how educational policy and teacher orientation affect their practices. In the following studies, teachers' demographic characteristics (gender, age, experience, school type etc.), attributes (e.g. motivation) and philosophy (e.g. curriculum orientation) should also be included as factors to understand the relationship of these factors with teachers' assessment practices and beliefs. Furthermore, in order to deepen our understanding on how teachers implement assessment in their classroom, the rationale behind their practices and the factors affecting their practices, comparative case studies should be conducted through not only using interviews but also using a series of observation (e.g. video study) and teacher documents.



References

- Aiken, L. R. (1997). *Questionnaires and inventories: Surveying opinions and assessing personality*. New York: John Wiley & Sons, Inc.
- Akbayrak, B. (2000). A comparison of two data collection methods: Interviews and questionnaires. *Hacettepe Üniversitesi, Eğitim Fakültesi Dergisi*, 18, 1-10.
- Alberta Education (2008). *Spanish language and culture nine-year program: Grade 4 classroom assessment material*. Alberta, Canada: Alberta Education, Learning and Teaching Resources. Retrieved on 09.02.2019 from files.eric.ed.gov/fulltext/ED506103.pdf.
- Al-Mahrooqi, R.I., & Denman, C.J. (2018). Alternative assessment. In J.I. Liontas (Ed.) *TESOL encyclopedia of English language teaching* (1st Ed.). John Wiley & Sons, Inc. (pp. 4851-4856). doi: 10.1002/9781118784235.eelt0325.
- Anandan, K. (2015). *Assessment for learning*. Bharathidasan University: Centre for Distance Education. Retrieved on 13.01.2019 from www.bdu.ac.in/cde/docs/ebooks/B-Ed/I/ASSESSMENT%20FOR%20LEARNING.pdf
- 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*, 74, Jossey-Bass Publishers
- Ayas, N. (2014). *An investigation of teachers' conception and practices concerning assessment in English for young learners classroom*. (Unpublished Master Thesis) Çağ University, Institute of Social Sciences, Mersin.
- Bachman, L. (1990). *Fundamental considerations in language testing*. MA: Oxford University Press
- Bachman, L., & Palmer A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

- Bachman, F. L. (2002). Alternative interpretations of alternative assessments: Some validity issues in educational performance assessments. *Educational Measurement: Issues and Practice*, 21(3), 5-18. doi: 10.1111/j.1745-3992.2002.tb00095.x
- Bailey, K. M. (1998). Learning about language assessment: dilemmas, decisions, and directions. Heinle & Heinle: US.
- Balliro, L. (1993). What kind of alternative? Examining alternative assessment. (*TESOL Quarterly*, 27(3), 558-561. doi: 10.2307/3587490
- Banaji, M. R., & Heiphetz, L. (2010). Attitudes. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.) *Handbook of social psychology* (pp. 348-388). New York, NY: John Wiley & Sons.
- Bhola, H.S. (1990) “*Literacy for Development*” projects, programs and campaigns: *Evaluation planning, design and implementation and utilization of evaluation results*. Hamburg, Germany: UNESCO Institute for Education; DSE. xii, p.306.
- Blood, D.F. & Budd, W.C.(1972). *Educational measurement and evaluation*. USA: Harper & Row Publishers.
- Britton, A.M. (2015). *Assessment for learning in teaching English to young learners: Teachers’ understanding, classroom practice and impact on interaction*. Unpublished Doctoral Dissertation, University of Reading. Retrieved on 10.02.2019 from centaur.reading.ac.uk/41939/1/21803102_Britton_thesis.pdf.
- Broadfoot, P. M. (22-25 July 2003). *Dark alleys and blind bends: Testing the language of learning*. Paper presented at the 25th Language Testing Research Colloquium, University of Reading.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education, *Assessment in Education: Principles, Policy & Practice*, 11(1), 7-26, doi: 10.1080/0969594042000208976

- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12.
doi:10.1111/j.1745-3992.2003.tb00139.x
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Longman
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science Publishers.
- Brown, H. D., & Abeywickrama, P. (2012). *Language assessment: Principles and classroom practices*. New York: Pearson Education.
- Brown, J. B., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653–75. doi: 10.2307/3587999
- 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, doi: 10.1080/03055690802648531
- Büyükkarcı, K. (2010). *The effect of formative assessment on learners' text anxiety and assessment preferences in EFL context*. Doctoral thesis, Çukurova University, the Institute of Social Science, Adana
- Büyükköztürk, Ş. (2005). Anket geliştirme. *Türk Eğitim Bilimleri Dergisi*, 3(2), 133-151.
- Cameron, L. (2001). *Teaching Language to Young Learners*. UK: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511733109>
- Chan, Y-C. (2008). Elementary school EFL teachers' beliefs and practice of multiple assessments. *Reflecting on English Language Teaching*, 7(1), 37-62.
- Clarke, S. (2005). *Formative assessment in action: Weaving the elements together*. London: Hodder Murray.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage Publication.

- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Conner, O. (2008). Testing as a catalyst for progress and development. *The International TEYL Journal*. Retrieved May 5, 2019 from <http://www.teyl.org/article7.html>
- Council of Europe (2001). *Common European Framework of References: Learning, teaching, assessment*. Strasbourg: Language Policy Unit. Retrieved on 16.02.2019 from www.coe.int/lang-CEFR.
- Culbertson, D. L. (2000). *Alternative assessment: Primary grade literacy teachers' knowledge, attitudes and practices*, Indiana University of Pennsylvania, Pennsylvania, United States.
- Cumming, A. (2009). Language assessment in education: Tests, curricula, and teaching. *Annual Review of Applied Linguistics*, 29, 90–100. doi: <https://doi.org/10.1017/S0267190509090084>
- Davison, C., & Cummins, J. (2007). Assessment and evaluation in ELT: Shifting paradigms and practices. In J. Cummins and C. Davison (Eds.), *International Handbook of English Language Teaching* (pp. 415 - 420). New York: Springer.
- Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *The Turkish Online Journal of Educational Technology*, 2(3), 13–19.
- Dochy, F. (2001). A new era: Different needs, new challenges. *Research Dialogue in Learning and Instruction* 2 (2001), 11-20.
- Doğandere, İ. C. (2006). *Perspectives of young learners of English towards assessment: Formative or summative?*. Unpublished Master Thesis, Uludağ University, The Institute of Social Sciences, Bursa.
- Faez, F. (2011). English Education in Italy: Perceptions of teachers and professors of English. *Canadian and International Education*, 40(3), 31-44.

- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.
- Gattullo, F. (2000). Formative assessment in ELT primary (elementary) classrooms. An Italian case study. *Language Testing*, 17(2), 278-288.
doi:10.1177/026553220001700210
- Genesse, F. (2001). Evaluation. In R. Carter & D. Nunan (Eds). *Teaching English to speakers of other languages* (pp. 144-150). Cambridge: Cambridge University Press.
- Genesee, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education*. New York: Cambridge University Press.
- Gonzales, J. A. 2009. Promoting students autonomy through the use of the European language portfolio. *ELT Journal*, 63(4), 373-382.
- Gökçe, Ö.F. (2014). *A comparison of EFL teachers' perception of formative assessment in public and private schools*. Unpublished Master Thesis, Çağ University, Institute of Social Sciences, Mersin.
- Griffiths, T. , Gore, J., & Ladwig, J. (2006, November). *Teachers' fundamental beliefs, commitment to reform and the quality of pedagogy*. Paper presented at Proceedings Australian Association for Research in Education Annual Conference, Adelaide.
- Gronlund, N. E., & Linn, R. L. (1985). *Measurement and evaluation in teaching* (6th ed.). MacMillan
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, 8(3), 381 – 391. doi: 10.1080/135406002100000512
- Halali, A.A.S., Singh, H.K.D., Saeed, I M., & Making Z. (2017). Teachers' perceptions toward the use of classroom-based assessment in English reading. *International Journal of Education and Research*, 5(11), 153-168.

- Hales, W. & Marshal, J.C. (1972). *Essentials of testing*. USA: Addison-Wesley Publishing Company.
- Hambleton, K. R., & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education*, 5(1), 1-16. doi: 10.1207/s15324818ame0501_1
- Harmer, J. (2001). *The practice of English language teaching*. Harlow: Longman.
- Hayes, D. (2014). *Factors influencing success in teaching English in state primary schools*. London: British Council. Retrieved on 17.02.2019 from www.teachingenglish.org.uk
- Heaton, J. B. (1990). *Writing English language test*. NY: Longman.
- Helminen, L. (18 November 2013). *Aspects of language teaching and training – examples of Finland*. Workshop for language teacher trainers in Budapest. Retrieved on 16.02.2019 from http://oktataskesztes.tka.hu/documents/Projekt/2013/TANDEM/WP5_nyelvtanarkepzes/aspects_of_languageteachingandtraining_hu.pdf
- Hobson, A, & Townsend, A. (2010). Interviewing as educational research methods. In D Hartas (Ed.), *Educational research and inquiry, continuum*. (pp. 223-238). London: Continuum.
- Hoffman, J. V., L. C. Assaf, & S. G. Paris. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher* 54(5): 482–92.
- Janisch, C., Liu, X., & Akrofi, A. (2007). Implementing alternative assessment: Opportunities and obstacles. *The Educational Forum*, 71, 221-229. doi: 10.1080/00131720709335007
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *Educational Research Association The International Journal of Research in Teacher Education*, 6(2), 26-37.

- Jia, Y., Eslami, Z.R., & Burlbaw, L.M. (2006). ESL teachers' perceptions and factors influencing their use of classroom-based reading assessment. *Bilingual Research Journal*, 30(2), 407 - 430. doi: 10.1080/15235882.2006.10162883
- Johnson, K. E (1992). The relationship between teachers' beliefs and practices during literacy instruction for non-native speakers of English. *Journal of Reading Behavior*, 24(1), 83–108. doi: 10.1080/10862969209547763
- Ioannou - Georgiou, S. (2003). *Assessing young learners (Resource books for teachers)*. Oxford: OUP.
- Kibar, E.P. (2018). *Pre-service and in-service English teachers perceptions of testing and assessment in EFL classes*. Unpublished Master Thesis, Hacettepe University, Institute of Educational Sciences, Ankara.
- Kırkgöz, Y. and Ağçam, R. (2012). Investigating the written assessment practices of Turkish teachers of English at primary education. *The Journal of Language and Linguistic Studies*, 8 (2), 119-136. Retrieved on 09.02.2019 from <http://www.jlls.org/vol8no2/119-136.pdf>
- Kırkgöz, Y., Babanoğlu, M.P., & Ağçam, R. (2017). Turkish EFL teachers' perceptions and practices of foreign language assessment in primary education. *Journal of Education and e-Learning Research*, 4(4), 163-170.
- Kohonen, V. (1997). Authentic assessment as an integration of language learning, teaching, evaluation and the teacher's professional growth. In A. Huhta, V. Cohonen, L. Kurki-Suonio and S. Luoma (Eds.), *Current Developments and Alternatives in Language Assessment: Proceedings of LTRC 1996* (pp. 7-22). University of Jyväskylä: Jyväskylä.
- Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. Thousands Oaks, CA: Sage.

- Lamprianou, I., & Athanasou, J.A. (2009). *A teacher's guide to educational assessment*. Rotterdam: Sense Publishers.
- Law, B., & Eckes, M. (1995). *Assessment and ESL*. Peguis publishers: Manitoba, Canada.
- Lee, A. S., Park, S. H., & Choi, J. (2012). The relationship between communication climate and elementary teachers' beliefs about alternative assessment. *Journal of Pacific Rim Psychology*, 5(1), 11-18. doi: <https://doi.org/10.1375/prp.5.1.11>
- Liljedahl, P. (2010). The four purposes of assessment. *Vector* (Summer), 7-12. Retrieved on 15.01.2019 from www.peterliljedahl.com/wp.../4-Purposes-of-Assessment.pdf
- Lopriore, L. (2002). The teaching of EFL in the Italian context: Issues and implications. *CAUCE, Revista de Filologia y suDidactica*, 25, 203-223.
- Luoma, S., & Tarnanen, M. (2003). Creating a self-rating instrument for L2 writing: From idea to implementation. *Language Testing*, 20(4), 440-465. doi: 10.1191/0265532203lt267oa
- McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- MEB – Milli Eğitim Bakanlığı (2018). *İngilizce dersi öğretim programı (İlkokul ve Ortaokul 2, 3, 4, 5, 6, 7 ve 8. Sınıflar)*. Ankara: Milli Eğitim Bakanlığı.
- Mehrens, W.A., & Lehmann, I.J. (1973). *Measurement and evaluation in education and psychology*. Rinehart and Winston Inc. USA.
- Mertler, C. A. (1998, October). *Classroom assessment practices of Ohio teachers*. Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago, IL. Retrieved on 06.02.2019 from <https://files.eric.ed.gov/fulltext/ED428085.pdf>

- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
doi: 10.1177/026553229601300302
- Nasab, F.G. (2015). Alternative versus traditional assessment. *Journal of Applied Linguistics and Language Research*, 2(6), 165-178.
- Nunan, D. (2011). *Teaching English to young learners*. USA: Anaheim University Press.
- Önalın, O., & Karagül, A.E. (2018). A study on Turkish EFL teachers' beliefs about assessment and its different uses in teaching English. *Journal of Language and Linguistic Studies*, 14(3), 190-201.
- Öz, H. (2014). Turkish teachers' practices of assessment for learning in the English as a foreign language classroom. *Journal of Language Teaching and Research*, 5(4), 775-785.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62 (3), 307-332. doi: 10.2307/1170741
- Palomba, C. & Banta, T.W. (1999) *Assessment essentials: Planning, implementing, and improving assessment in higher education*. Jossey-Bass Publishers: San Francisco.
- Pinter, A. (2006). *Teaching young language learners*. Oxford, United Kingdom: Oxford University Press
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429-462. doi: 10.1177/026553220101800407
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing* 21(3), 249-258. doi: 10.1191/0265532204lt283ed
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472-482. doi: 10.1016/j.tate.2010.09.017

- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.
- Şahinkarakaş, S. (2012). The role of teaching experience on teachers' perceptions of language assessment. *Procedia – Social and Behavioral Sciences* 47, 1787-1792.
- Saito, H., & Inoi, S. I. (2017). Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly*, 14(3), 213-233. doi: 10.1080/15434303.2017.1351975
- Shim, K. N. (2009). *An investigation into teachers' perceptions of classroom-based assessment of English as a foreign language in Korean primary education*. Unpublished Doctoral Dissertation, University of Exeter, Exeter. Retrieved on 08.02.2019 from <https://ore.exeter.ac.uk/repository/bitstream/handle/10036/.../ShimKN.doc..pdf?...2>
- Smith, K. (1996). Assessing and testing young learners: Can we? Should we? In D. Allen (Ed.). *A Symposium of the research, testing, and young learners special interest groups* (pp.1-9). Kent, England: IATEFL.
- Song, E., & Koh, K. (22–27 August 2010). *Assessment for learning: Understanding teachers' beliefs and practices*. Paper presented at the 36th International Association for Educational Assessment (IAEA) Annual Conference in Bangkok, Thailand. Retrieved on 06.02.2019 from <https://repository.nie.edu.sg/handle/10497/16733>
- Şentürk, F. (2013). *Washback effect of ket exam in learning English as a foreign language*. Unpublished Master Thesis, Çağ University, Institute of Social Sciences, Mersin.
- Tatlı, L. (2014). *EFL teachers' attitudes toward teaching English to young learners*. Unpublished Master Thesis, Çağ University, Institute of Social Sciences, Mersin.
- Terrell, S. (2011). Mixed-methods research methodologies. *The Qualitative Report*, 17(1), 254-280. Retrieved from <http://www.nova.edu/ssss/QR/QR17-1/terrell.pdf>

- Thomas, R. M. (1998). *Conducting educational research: A comparative view*. West Port, Conn: Bergin & Garvey.
- Thomas, M. (2012). Teachers' beliefs about classroom assessment and their selection of classroom assessment strategies. *Journal of Research and Reflections in Education*, 6(2), 103 -112
- Veal, M. L. (1988). Pupil assessment perceptions and practices of secondary teachers. *Journal of Teaching in Physical Education*, 7 (4), 327-42. doi: [10.1123/jtpe.7.4.327](https://doi.org/10.1123/jtpe.7.4.327)
- Yıldırım, 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.
- Wewer, T. (2014). *Assessment of young learners' English proficiency in bilingual content instruction CLIL*. Unpublished Doctoral Dissertation, University of Turku, Turku, Finland.
- Wiesnerová, D. (2012). *Benefits of self-assessment in English classes at elementary schools*. Bachelor Thesis, Masaryk University Brno. Retrieved on 15.01.2019 from https://is.muni.cz/th/h8x8n/Bachelor_Thesis_Wiesnerova.pdf
- Zorba, M. G. and Tosun, S. (2011). Enriching kindergarten learners' English by using language portfolio and additional instructional materials. *Contemporary Online Language Education Journal*, 2011, 1(2), 35-43.

APPENDICES

Appendix 1: Questionnaire

Dear Colleague,

This questionnaire is designed to identify teachers' practices and beliefs about the assessment of 4th grade EFL students in Turkey and some European Countries for MA in ELT thesis.

Please tick (✓) the square () before the appropriate answer or write your own answer. We appreciate your cooperation in completing this questionnaire.

Thank you for your participation.

English Language Teacher Nilay ÇAKIR
Advisor: Assoc. Prof. Dr. Zübeyde Sinem GENC

PART A. CHARACTERISTICS OF TEACHERS' ASSESSMENT IN PRACTICE

1. Gender:

- Male Female

2. Country:

- Turkey Italy Finland

3. I have been teaching English for ...

- 0 to 5 years 6 to 15 years 16 to 24 years 25 years or more

4. To teach English I have

- Bachelor's Degree (Lisans) Master's Degree (Yüksek Lisans)

5. I work in a ...

- public school private school

6. Typical class size is:

- Less than 20 students 20 to 29 students 29 to 35 students More than 35 students

7. Do you assess your students in any way on their achievements at 4th Grade ?

- Yes No

8. Why is assessment carried out in your classroom? (Birden fazla seçenek işaretleyebilirsiniz)

- To check students strength and weaknesses for lesson planning

- To provide information to parents and students
- To place students
- To motivate students to learn
- To provide a record of students' achievement for success and failure

Other, please specify:

9. Who is responsible for the preparation of assessment? (Birden fazla seçenek işaretleyebilirsiniz)

- A foreign language teacher
- A primary teacher
- A group of language teachers within the school
- Local education authority
- National education authority

10. How are the exam results reported? (birden fazla seçenek işaretleyebilirsiniz)

- Numerical grade
- Comments
- Grade and comments

Other, please specify:

11. Before the exam, do you inform your students about the criteria of your assessment?

- Yes No

12. Are you additionally trained to prepare and perform assessment of students in primary schools?

- Yes No

13. If Yes, by whom? (Birden fazla seçenek işaretleyebilirsiniz)

- at university
- in seminars organized by the Ministry of Education

from colleagues

in ELT seminars and conferences

Other, please specify:

14. What type of assessment do you use in your FL classroom? 4.sınıflarda hangi tür sınav kullanıyorsunuz?

(Her madde için cevaplardan yalnızca birini işaretleyiniz)

Types of Assessment	Answers				
	(Always: once a week; Frequently: once a fortnight; Sometimes:2–3 times a term; Rarely: once a term)				
My own-made paper and pencil test	Always	Frequently	Sometimes	Rarely	Never
Standardized paper and pencil test from national or local EFL education authority	Always	Frequently	Sometimes	Rarely	Never
Performance-based test (eg. Role-play, drama..)	Always	Frequently	Sometimes	Rarely	Never
Structured observation leading to a written description of students' performance	Always	Frequently	Sometimes	Rarely	Never
Oral interview	Always	Frequently	Sometimes	Rarely	Never
Language portfolio	Always	Frequently	Sometimes	Rarely	Never
Self-assessment	Always	Frequently	Sometimes	Rarely	Never
Peer-assessment	Always	Frequently	Sometimes	Rarely	Never

15. What is your content/skills-focus in assessment of students? (Sınavlarda hangi yeterlilik ve becerileri ölçüyorsunuz?)

(Her madde için cevaplardan yalnızca birini işaretleyiniz)

Content and skills	Answers				
	Always	Frequently	Sometimes	Rarely	Never
Grammar	Always	Frequently	Sometimes	Rarely	Never
Lexis (Vocabulary)	Always	Frequently	Sometimes	Rarely	Never
Spelling	Always	Frequently	Sometimes	Rarely	Never
Speaking skills	Always	Frequently	Sometimes	Rarely	Never
Listening Skills	Always	Frequently	Sometimes	Rarely	Never
Reading Skills	Always	Frequently	Sometimes	Rarely	Never
Writing Skills	Always	Frequently	Sometimes	Rarely	Never
Pronunciation	Always	Frequently	Sometimes	Rarely	Never

16. What kind of exercises do you use in your assessment? (Sınavlarınızda hangi tür egzersizler kullanıyorsunuz?)

(Her madde için cevaplardan yalnızca birini işaretleyiniz)

Exercises	Answers				
	Always	Frequently	Sometimes	Rarely	Never
Filling gaps	Always	Frequently	Sometimes	Rarely	Never
Matching	Always	Frequently	Sometimes	Rarely	Never
Role-playing	Always	Frequently	Sometimes	Rarely	Never
Writing words, short texts	Always	Frequently	Sometimes	Rarely	Never
Comprehension questions (e.g. true/false) about a short text	Always	Frequently	Sometimes	Rarely	Never
Dictation	Always	Frequently	Sometimes	Rarely	Never
Grammatical activities	Always	Frequently	Sometimes	Rarely	Never
Students create their own sentences/texts	Always	Frequently	Sometimes	Rarely	Never
Listening to audio material (e.g. short films, clips)	Always	Frequently	Sometimes	Rarely	Never

Making a dialogue	Always	Frequently	Sometimes	Rarely	Never
Students repeat vocabulary or sentences	Always	Frequently	Sometimes	Rarely	Never
Multiple Choice	Always	Frequently	Sometimes	Rarely	Never
Reading words and sentences	Always	Frequently	Sometimes	Rarely	Never
Oral description of a picture	Always	Frequently	Sometimes	Rarely	Never
Spelling	Always	Frequently	Sometimes	Rarely	Never

PART B. TEACHERS' BELIEFS ABOUT ASSESSMENT

1. In your opinion, what type of assessment should be used for 4th grade students?

(4.sınıflarda hangi tür sınav kullanılmalı?)

TYPES	I fully agree	I agree	I partly agree	I don't agree	I don't agree at all
My own-made paper and pencil test					
Standardized paper and pencil test from national or local EFL education authority					
Performance-based test (eg. Role-play, drama..)					
Structured observation leading to a written description of students' performance					
Oral interview					
Language portfolio					
Self-assessment					
Peer-assessment					

2. In your opinion, what should be the content/skills-focus in assessment for 4th grade students? (Sınavlarda hangi yeterlilik ve becerilerin ölçülmesi gerektiğini düşünüyorsunuz?)

TYPES	I fully agree	I agree	I partly agree	I don't agree	I don't agree at all
Grammar					
Lexis (Vocabulary)					
Spelling					
Speaking skills					
Listening Skills					
Reading Skills					
Writing Skills					
Pronunciation					

3. In your opinion, what kind of exercises should be used for 4th grade students?

(Sınavlarda hangi tür egzersizler kullanılmalı?)

Statements	I fully agree	I agree	I partly agree	I don't agree	I don't agree at all
Filling gaps					
Matching					
Role-playing					
Writing words, short texts					
Comprehension questions (e.g. true/false) about a short text					
Dictation					
Grammatical activities					
Students create their own sentences/texts					
Listening to audio material (e.g. short films, clips)					
Making a dialogue					
Students repeat vocabulary or sentences					
Multiple Choice					
Reading words and sentences					

Oral description of a picture					
Spelling					

4. Confirm or deny the following statements about assessment of students' (SS) foreign language.

Instruction: Please add the sign (✓) to the individual statement in the suitable column

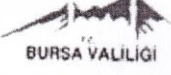


Statements	I fully agree	I agree	I partly agree	I don't agree	I don't agree at all
Students of a foreign language should be assessed in the primary school.					
I need to be more trained in assessing students in primary schools.					
It is difficult to assess speaking skills in the crowded classes.					
There should be ready exams in the books to assess listening skills.					
Students should be assessed to check students strength and weaknesses for lesson planning					
Students should be assessed by comments, not by numerical grades.					
The results of assessment should be used to place students.					
There should be ready exams in the books to assess speaking skills.					
Students should be assessed to provide a record for success and failure.					
Students should be tested on sentence-level activities.					
Parents should always be informed with the results of students' assessment.					

Appendix 2: Semi-Structured Interview Questions

Interview Questions

1. What do you think about the exams that you use for your young learners in your classroom?
Are the types or exercises of the exams appropriate for the students? Do they evaluate the right skills?
2. How do you assess communicative skills (listening, speaking etc.)? Do you think that they are effective?
3. Would you use exams that assess communicative skills more if you found them ready in the course books or on an online platform?
4. Do you think using language portfolio as an assessment in your FL classroom is useful and necessary? Why? /Why not?
5. What do you think about using self or peer evaluation as an assessment? Do they have positive or negative effects on students?
6. What are the difficulties that you have in assessing young learners ?
7. Do you have any solutions for these problems?
8. What type of assessment would you use if you had proper conditions in your classroom?
(What is the best exam according to you?)
9. How can you improve your assessing skills?
10. Do you have any more suggestions or comments about assessing young learners?

Appendix 3: Bursa Provincial Directorate of National Education Approval Letter

 <p>BURSA VALİLİĞİ</p>	 <p>T.C. BURSA VALİLİĞİ İl Millî Eğitim Müdürlüğü</p>	 <p>BURSA</p>
<p>Sayı : 86896125-605.01-E.6388422 Konu : Nilay CEYLAN'ın Araştırma İzni</p>	<p>28.03.2018</p>	
<p>MÜDÜRLÜK MAKAMINA</p>		
<p>İlgi : Millî Eğitim Bakanlığı'nın Araştırma, Yarışma ve Sosyal Etkinlik İzinleri konulu 22/08/2017 tarihli ve 2017/25 sayılı Genelgesi.</p>		
<p>Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü Yabancı Diller Eğitimi Anabilim Dalı İngiliz Dili Eğitimi Bilim Dalı yüksek lisans programı öğrencisi Nilay CEYLAN'ın "Türkiye ve İki Avrupa Ülkesindeki öğretmenlerin 4. Sınıf EFL Öğrencilerinin Değerlendirmelerine Yönelik Görüş ve Uygulamalarının Karşılaştırmalı Analizi" konulu araştırma isteği Uludağ Üniversitesi Rektörlüğü Genel Sekreterlik'in 16/03/2018 tarihli ve 9808 sayılı yazısı ile bildirilmektedir.</p>		
<p>Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü Yabancı Diller Eğitimi Anabilim Dalı İngiliz Dili Eğitimi Bilim Dalı yüksek lisans programı öğrencisi Nilay CEYLAN'ın "Türkiye ve İki Avrupa Ülkesindeki öğretmenlerin 4. Sınıf EFL Öğrencilerinin Değerlendirmelerine Yönelik Görüş ve Uygulamalarının Karşılaştırmalı Analizi" konulu araştırmasını Müdürlüğümüze bağlı ilkokullarda görevli İngilizce öğretmenlerine online araştırma yapma isteği ilimizde oluşturulan "Araştırma Değerlendirme Komisyonu" tarafından incelenerek değerlendirilmiştir. Araştırma ile ilgili çalışmanın okul/kurumlardaki eğitim öğretim faaliyetleri aksatılmadan, araştırma formlarının aşı okul müdürlüklerince görülerek ve gönüllülük esası ile okul müdürlüklerinin gözetim ve sorumluluğunda ilgi Genelge çerçevesinde uygulanması ayrıca araştırma sonuçlarının Müdürlüğümüz ile paylaşılması komisyonumuzca uygun görülmektedir.</p>		
<p>Makamlarınızca da uygun görülmesi halinde olurlarınıza arz ederim.</p>		
<p>Müştak GENCER İl Millî Eğitim Şube Müdürü</p>		
<p>OLUR 28.03.2018</p>		
<p>Mustafa KAHYA Vali a. İl Millî Eğitim Müdür V.</p>		
<p>Adres : Hocahasan Mh. İlkbahar Cad. No:38 (Yeni Hükümet Konakı A Blok) 16050-Osmangazi/BURSA Telefon No:(0224) 445 16 00 Fax: 445 18 10</p>	<p>Bilgi İçin : Ekrem KOZ İl Millî Eğitim Müdür Yardımcısı Tel: (0224) 445 1638</p>	<p>Leyla DİKİCİ VHKİ (0224) 215 25 39</p>

Appendix 4: Uludağ University Research Ethics Committee Approval Letter



T.C.
ULUDAĞ ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü Müdürlüğü



Sayı: 20585590-806.01.03/242

27/01/2017

Konu: Nilay CEYLAN'ın Tez Konusu Önerisi

YABANCI DİLLER EĞİTİMİ ANABİLİM DALI BAŞKANLIĞINA

İlgi : 19.01.2017 tarihli ve 41401206-806.01.03/164 sayılı yazınız.

Enstitümüz Yönetim Kurulunun 20.01.2017 tarih ve 2017/02/02 sayılı kararı ile Anabilim Dalınız Öğretim Üyesi Doç.Dr.Zübeyde Sinem GENÇ'in danışmanı olduğu Yüksek Lisans öğrencisi Nilay CEYLAN'ın "Türkiye ve İki Avrupa Ülkesindeki Öğretmenlerin, 4. Sınıf EFL Öğrencilerinin Değerlendirmelerine Yönelik Görüş ve Uygulamalarının Karşılaştırmalı Analizi." tez konusunun uygun olduğuna karar verilmiştir.

Bilgilerinizi ve konunun Danışman Öğretim Üyesine bildirilmesi konusunda gereğini rica ederim.

imza

Doç. Dr. Esim GÜRSOY
Müdür V.

U.Ü. Eğitim Bilimleri Enstitüsü Görükle Kampusu 16059 Nilüfer/BURSA

Tel : 0224 2940975 Faks: 0224 2940975

e-posta : egtbil@uludag.edu.tr

Elektronik Ağ: www.uludag.edu.tr

Bilgi İçin: Ali ALKIŞ

Memur

Tel : 0224 2942883

Bu belge UDOS ile hazırlanmıştır. Teyit için: <https://udos.uludag.edu.tr/teyit/?eD6ex9VohUCY700vQCjTVA>



Öz Geçmiş

Doğum Yeri ve Yılı : Bursa - 1987

Öğr. Gördüğü Kurumlar	Başlama Yılı	Bitirme Yılı	Kurum Adı
Lise	2001	2005	Bursa Kız Lİsesi
Lisans	2005	2009	Bursa Uludağ Üniversitesi
Lisans	2015	2019	Anadolu Üniversitesi
Yüksek Lisans	2015	2019	Bursa Uludağ Üniversitesi

Bildiği Yabancı Diller ve Düzeyi : İngilizce – İleri

Çalıştığı Kurumlar	Başlama ve Ayrılma Tarihleri	Kurum Adı
	1. 2010 - 2010	Namık Kemal İlköğretim Okulu
	2. 2010 - 2013	Baraklı İlköğretim Okulu
	3. 2013 – 2015	Ukla Akademi Dil Okulu
	4. 2015- 2015	Bursa İl Milli Eğitim Müdürlüğü
	5. 2013- 2018	Şehit Gürcan Ulucan İlkokulu
	6. 2018 – 2019	Kadir Koyutürk İlkokulu
	7. 2019 – Halen	İzmit İlçe Milli Eğitim Müdürlüğü

Yurt Dışı Görevleri : -

Kullandığı Burslar : -

Aldığı Ödüller : -

Üye Olduğu Bilimsel ve Mesleki Topluluklar : -

Editör veya Yayın Kurulu Üyeliği : -

**Yurt İçi ve Yurt Dışında
Katıldığı Projeler :**

- 1.Rural Rules-Youth exchange project in Romania (10.10.2014)
- 2.More than Money: Social Entrepreneurship Heroes -Youth exchange project in Lituania (05.12.2014)
- 3.REACT -Active Citizenship- Youth exchange project in Portugal (26.01.2015)
- 4.SESAME- Ka2 Strategic partnership transnational meeting in UK (15.06.2015)
- 5.IQ-Involment to Quality- Youth exchange project (11.09.2015)
- 6.EVS for Employability - Training Course (22.11.2015)
7. Turning mirrors into windows -Training Course (13.04.2016)
8. All in the Same Boat - Becoming an Activist Against Hate-Training Course in Portugal (28.02.2017)

**Katıldığı Yurt İçi ve Yurt
Dışı Bilimsel Toplantılar : -****Yayımlanan Çalışmalar :**