HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Department of Foreign Languages Education

English Language Teaching Program

ASSESSING WRITING IN ACCORDANCE WITH THE DESCRIPTORS OF THE
COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES

Hamdiye AVCI

A Master's Thesis

Ankara, 2019

With leadership, research, innovation, high quality education and change,

To the leading edge… Toward being the best…

HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Department of Foreign Languages Education

English Language Teaching Program

ASSESSING WRITING IN ACCORDANCE WITH THE DESCRIPTORS OF THE COMMON EUROPEAN FRAMEWORK OF REFERENCE FOR LANGUAGES

YAZMA BECERİSİNİN AVRUPA DİLLERİ ÖĞRETİMİ ORTAK ÇERÇEVE PROGRAMI (CEFR) BETİMLEYİCİLERİ DOĞRULTUSUNDA DEĞERLENDİRİLMESİ

Hamdiye AVCI

A Master's Thesis

Ankara, 2019

Acceptance and Approval

To the Graduate School of Educational Sciences, This thesis / dissertation, prepared by HAMDİYE AVCI and entitled "**Assessing Writing in Accordance with the Descriptors of the Common European Framework of Reference for Languages**" has been approved as a thesis for the Degree of Master in the Program of English Language Teaching in the Department of Foreign Languages Education by the members of the Examining Committee.

Chair   Assoc. Prof. Dr. Hacer Hande UYSAL

Member (Supervisor)   Assist. Prof. Dr. Hatice ERGÜL

Member   Assist. Prof. Dr. Betül BAL GEZEGİN

This is to certify that this thesis/dissertation has been approved by the aforementioned examining committee members on 22/08/2019 in accordance with the relevant articles of the Rules and Regulations of Hacettepe University Graduate School of Educational Sciences, and was accepted as a Master's Thesis Dissertation in the Program of English Language Teaching by the Board of Directors of the Graduate School of Educational Sciences on ..../......./.........

Prof. Dr. Ali Ekber ŞAHİN
Director of Graduate School of Educational Sciences

Abstract

The purpose of this thesis is to analyse the writing tasks of A2 level students in accordance with the CEFR descriptors at Karabuk University, School of Foreign Languages. To that end, the writing tasks of the focus group, which had been appointed as 93 A2 level students before, were studied to find out whether they had been assessed based on a CEFR-based rubric. Two different writing tasks of the focus group students were graded by the instructors. The tasks were graded by the standardization team, too and the consistency between the scores given by the instructors and the standardization team was evaluated. An A2 level "Can Do Statements" questionnaire approved by the Council of Europe was conducted so as to get a better profile of the focus group. The data were analyzed. Moreover, an interview with the instructors of the focus group was carried out to evaluate the CEFR practices they had been carrying out with respect to not just students, but instructors as well. The data was gathered as written documents, and then evaluated. The results indicated that there was not a significant difference between the writing task scores given by the instructors and the standardization team. Likewise, the findings displayed that instructors highly valued the CEFR practices and they also found them helpful. On the other hand, the rubric that was used by the instructors while grading tasks was found to be a little different from the CEFR-based one.

Keywords: common european framework of reference for languages (cefr), assessment, descriptors, criteria, writing.

Öz

Bu çalışmanın amacı, Karabük Üniversitesi Yabancı Diller Yüksekokulu A2 düzeyi öğrencilerinin yazma becerilerinin Avrupa Dilleri Öğretimi Ortak Çerçeve Programı ölçütlerine göre analiz etmektir. Bu amaçla, 93 kişi olarak belirlenen odak grubu öğrencilerinin yazma çalışmalarının CEFR kriterleri baz alınarak değerlendirilip değerlendirilmediği araştırıldı. Belirlenen grubun öğretim görevlileri 2 ayrı yazma çalışmasını kullandıkları ölçekle notlandırdı ve bu sonuçlar standardizasyon ekibince de değerlendirilerek öğretmenler ve standardizasyon ekibinin uyumu değerlendirildi. Belirlenen öğrenci grubunun profilini çıkarmak amacıyla Avrupa Konseyi'nce onaylanmış olan Avrupa Dil Portfolyosu A2 düzeyi "Yapabiliyorum' ifadelerinden oluşan bir anket uygulanmıştır. Elde edilen veriler değerlendirilmiştir. Bunun yanı sıra, çalışmanın sadece öğrenciler değil öğretmenler açısından da değerlendirilmesi için belirlenen grubun öğretmenleriyle CEFR ile ilgili çalışmaları hakkında yüzyüze görüşme sağlanmış, veriler yazılı olarak toplanmış ve değerlendirilmiştir. Sonuçlar öğretim görevlileri ve standardizasyon ekibi tarafından değerlendirilen yazma çalışmalarının sonuçları arasında anlamlı bir farklılık olmadığını gösterdi. Sonuçlar ayrıca, öğretim görevlilerinin CEFR çalışmalarına oldukça önem verdiğini ve bu çalışmaları yararlı bulduklarını göstermiştir. Diğer yandan öğretim görevlileri tarafından kullanılan rubriğin CEFR-temelli olandan oldukça olduğu belirlenmiştir.

Anahtar sözcükler: avrupa dilleri öğretimi ortak çerçeve programı, değerlendirme, betimleyiciler, ölçütler, yazma becerisi.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ALTE   : The Association of Language Testers in Europe

ASK   : A Language Learner Corpus of Norwegian as a Second Language

CEFLING : The Linguistic Basis of the CEFR Levels: Combining Second

                Language Acquisition and Language Testing Research

CEFR   : Common European Framework of Reference

COE   : Council of Europe

DIALANG : European Commission-funded Project, through Socrates Lingua

                Action D to Create a Battery of On-line Diagnostic Tests in 14

                European Languages

EFL   : English as a Foreign Language

ELP   : The European Language Portfolio

EU   : European Union

SFL   : School of Foreign Languages

SPSS   : Statistical Package for Social Sciences

Chapter 1

Introduction

The idea of having a greater unity among European Countries, the member states of the Council of Europe introduced a foreign language education guideline based on the European plurilingual and multicultural citizenship identity policy. The guideline was called "The Common European Framework of Reference for Languages: Learning, Teaching, Assessment", and its publication in 2001 was the outcome of countless meetings and studies over a decade or so. The CEFR aims at promoting co-operation among educational institutions serving in different member countries of European Union so as to create a unique and sound basis for the mutual understanding of language features as well as to help learners, teachers, school administrators and coordinators reach a common basis in syllabuses, school curricula, examinations and course designs. As there are various educational systems in Europe, the CEFR has the intention of resolving the difficulties on communication and collaboration among the professionals working in these educational systems (Council of Europe, 2001).

The CEFR aims to standardize the levels of proficiency that are called Common Reference Levels, which enable teachers to measure learners' progress at each and every phase of learning. There are well-defined and standardized expectations in can-do statements form at each of these levels, which clearly state what is required from learners. What teachers and learners should do is to design appropriate activities to reach the levels and meet the required standards .

Providing a solid basis for the explicit description of content, methods and targets, the CEFR betters the transparency of syllabuses and courses. The objective criteria feature of the framework to determine language proficiency will meticulously help facilitate the common recognition of features gained in various contexts, and as a consequence of this, European mobility will pick up (Council of Europe, 1997).

This chapter propounds the background of the study, statement of the problem, the purpose of the study, research questions, method, hypothesis, the scope of the study, limitations, definitions and the conclusion part of this chapter.

Background to the Study

The Common European Framework of Reference for Languages (CEFR) studies on systematizing and standardising teaching and learning of modern languages in Europe as a part of the Council of Europe (COE) incepted 40 years ago. These studies basically aimed to serve "The Council of Europe's key political aims: the preservation of linguistic and cultural diversity, the promotion of linguistic and cultural tolerance, the promotion of plurilingualism, and education for democratic citizenship" (Little & Perclova, 2001, p. 3). What makes the CEFR a powerful guide is that it serves as a bounding source which emphasizes learner-centered teaching, and fosters standardisation and also learner autonomy among the 25 member countries of the EU (Council of Europe, 2001).

Turkey has 175 (71 private and 104 state) universities in total (ÖSYM, 2018). Out of 175, 90 universities have Faculties of Education, and all those faculties in Turkey are expected to act in accordance with a standardized curriculum formed by the Higher Education Council (Deniz & Sahin, 2006). On the other hand, the integration of Turkey into the EU requires systematic changes in education system, not to mention the criteria necessitating a long and demanding process that obliges improvement and adjustments in education. As an attempt to meet the requirements stated by the EU, the Ministry of National Education requires the use of the CEFR which is adopted as the source promoting standardized teaching and learning objectives, methods and assessment tools, and to this end, pilot studies were started at private high schools in 2002 (Kalkan, 2017).

Turkey aims to create a compatible higher education across EU member countries, so language teaching and learning practices in EU member countries have been adopted in our country as well (Mirici, 2017). This study aims to serve this purpose by shedding light on the assessment of writing tasks of students that they take as writing exams, taking into consideration the CEFR descriptors and analysing how much these descriptors help get a standardised writing assessment system when applied.

Statement of the Problem

The CEFR at all levels sets out the desired subskills on account of "can do" statements (see Appendix C), which help curriculum designers, teachers and learners make sure whether the CEFR levels are attained and learning criteria appropriate to the qualifications are included. The appropriate learning criteria, activities, tasks and assessment types should be considered accordingly. In this sense, the CEFR is regarded as an essential tool to provide the desired unity in language proficiency in educational institutions.

In the progress report of European Commission in 2016, it is stated that compared to the member countries of EU, foreign language teaching and learning studies do not seem enough in Turkey, and Turkey is reported as "partially ready in terms of education" (European Commission, 2016, p. 95). In Bayraktaroglu's report in 2014, on the other hand, it is pointed out that the main problem concerning foreign language education at universities is that accumulation of theoretical and practical knowledge along with the unity needed for a standardized education system is inadequate (Bayraktaroglu, 2014), which makes the CEFR practices necessary.

With the aim of analysing to what extent CEFR practices are applied at university level and how successful they are at solving the problem stated by Bayraktaroğlu (2014) and European Commission (2016) above, one of the universities that has been following the CEFR principles in their practices was needed to be researched into, so Karabuk University was considered to be an appropriate sample for it.

Karabuk University, the School of Foreign Languages Department (SFL), primarily adopted the CEFR as the guiding source in its education system in 2014, along with which the academic coordinators, skill coordinators and the administration staff have been putting great effort in designing the curriculum in accordance with the CEFR principles since then. As put by the vice principal of the university, one of the biggest issues has always been to have a standardised writing assessment. There are 70 instructors teaching at the SFL currently, and it is not an easy job to have all instructors grade the writing tasks of the students normatively, which necessitates to have a sound basis for writing assessment.

Starting from this point, a writing assessment system based on the CEFR descriptors was adopted, and this thesis aims to clarify if and to what extent this system has served the purpose.

As for the specifics about the exams, in an academic year at the SFL, there are three terms (three-month long in each term) and in each term; midterm, final, speaking and writing examinations are held. Both the learners and the instructors at SFL share some concerns on these assessment systems, specifically on the writing assessment as stated above. Given the circumstances, this research can be of service to have an acknowledged momentum in language teaching, learning and writing assessment.

Purpose of the Study

In todays' globalized world, teaching and learning foreign languages have become extremely important for intercultural communication. One of the most significant projects introduced with this aim is the CEFR by the Council of Europe, taking into consideration the needs and priorities of the member countries (Arslan & Coskun, 2012). Turkey has attempted to adapt the CEFR criteria to its education system and follow the process since the early 2000s. Yet still, have the required and necessary educational conditions for foreign language teaching and learning been prepared?

The idea to work on the CEFR and writing was the result of the researcher's 2-year-experience in testing unit and 1 year in standardisation committee at Karabuk University. One of the biggest issues that was faced with was the difference between the instructors' gradings of students' writing tasks and there always were meetings where that topic was discussed. After three years, the administration decided to adapt the CEFR to their system. The Framework presented a scaling system of six language proficiency levels: A1, A2, B1, B2, C1 and C2 (Council of Europe, 2001), can-do statements and descriptors, which were good points to start with. Starting from this point, a modular system was applied, so working on the CEFR and writing to see if the system solves the problem about the gradings, and to determine how effectively and to what extent the CEFR descriptors are used while assessing the writing tasks of the A2 Level students at Karabuk University, School of Foreign Languages seemed appropriate.

And why A2 Level? The language learners at A2 level are expected to be able to understand very basic sentences and expressions, communicate in simple tasks and use expressions so as to explain some features of his/her background (Council of Europe, 2001). As for written production and interaction skills, they are expected to be able to write simple phrases and sentences (about family, job, educational background..,etc.), very short descriptions, personal experiences, very simple letters expressing apology and thanks, short and simple notes, messages, explain likes, dislikes and opinions and write short and simple biographies (Council of Europe, 2001). In the modular system at Karabuk University, it was observed that A2 Level is the hardest level for the students to perform the skills stated above and pass the module.

On the other hand, In a report called 'English Language Teaching at Higher Education Institutions in Turkey' by British Council, it was stated that the subgrade level for the schools of foreign languages in Turkey should be at A2 Level as the students were supposed to perform the skills mentioned above by Council of Europe (West, Guven & Parry, 2015).

The main purpose of this study is to;

- use the CEFR as a tool for writing assessment.
- find out how different the CEFR (as a tool for writing assessment) is compared to the rubric used by the instructors.
- have two writing tasks written by the students and graded by the instructors depending on the rubric they have already been using and by the standardization team as well.
- see the differences and the similarities between the judgments of the instructors concerning the CEFR practices.

It is also necessary to see how much the instructors know about the CEFR. For this study, 4 instructors agreed to help gather data concerning their writing classes, and by means of an interview with these instructors, it is aimed to offer an insight into this need and have a better understanding about their perceptions on the CEFR.

On the other hand, in order to determine the students' levels and study from their point of view, they are provided with a list of 'can-do statements' with which they self-evaluate themselves and this serves as a measurement tool to understand their level.

The data gathered help the researcher further inform the instructors about the Common European Framework of Reference. It is also of great importance to help standardize the curriculum, assessment processes and syllabus.

Research Questions

The research was carried out with A2 (Waystage) level students and the instructors offering writing course to them. The rubric, "can do" statements and the interview with the instructors to be presented are expected to contribute both the students and the instructors on assessing writing at the A2 (Waystage) Level of the CEFR.

There are five research questions in the framework of this study:

1. Are the writing tasks of the A2 level prep-class students assessed in accordance with the CEFR descriptors?
   a) Are the rubrics that the instructors offering writing course use based on the CEFR descriptors?
2. To what extent do the instructors' gradings conform with the CEFR standardisation team's gradings?
3. What do the instructors offering writing course know about the CEFR?
4. What are the instructors' viewpoints on assessing writing in accordance with the CEFR descriptors?
5. What are the instructors' viewpoints on the contribution of the CEFR to teacher development?

Method

The research setting is Karabuk University, School of Foreign Languages. As it was stated in the purpose of the study part, Karabuk University decided to adapt the CEFR in its education system in 2014, and since then, great progress has been achieved. That is the main reason why this institution is a suitable sample to do research on the CEFR practices. As one of the biggest problem was

the gradings of the instructors on writing tasks, the main focus of the study, writing assessment, took form.

93 preparatory class A2 level students and 4 instructors offering writing course to this focus group took part in the study. The students' level was also determined by a European Language Portfolio (ELP) questionnaire. The learners had this chance by responding to "can do" statements that were to be given them at the very beginning of the study to determine that they really were at A2 level. Before the study, the focus group members were asked to fill in a consent form prepared to inform them about the process and to give them further information about the significance of the study, which was believed to motivate them to be a part of it and to want to be more enthusiastic. The tool is a quantitative one, so the responses of "can-do" statements questionnaire were analyzed via SPSS 22. The answers of 93 students for each statement were analyzed, and mean scores and standard deviations of them were calculated.

The CEFR descriptors, on the other hand, was used as a tool for writing assessment for the writing tasks of the students. The CEFR-based rubric was compared with the rubric being used by the instructors of the A2 level focus group. A comparative and thematic analysis was used to indicate the differences and similarities between the rubrics. The inferences of this study are believed to shed light upon the importance of writing assessment process, and the findings will be guiding for further studies.

As the qualitative phase of the study, a semi-structured interview with the instructors was carried out since they are considered as the most crucial element in the CEFR studies, and they serve the system as the course designers, the planners and the operators (Council of Europe, 2002). Their part in having a standardized, common language learning and teaching system cannot be underestimated, so the instructors involved in the study were asked for their opinions on the CEFR by answering a few questions. To analyze the data that were collected during 15 days, contextualising and categorising strategies were adopted. With the data gathered, it is believed that the problematic areas will be detected and the data will be used in further research studies. .

As the last part of the study, the grades by the instructors offering writing course and the standardisation team were compared and mean scores and standard deviations of them were calculated via SPSS 22.

Hypothesis

Since the CEFR is a crucial concept designed by Council of Europe for creating a sound basis for language teaching, Turkey needs to work on creating different course materials and assessment systems suitable for the six language proficiency levels described in the framework in order to keep pace with the EU countries (Arslan & Coskun, 2012). In this study, the main focus is the A2 (Waystage) proficiency level, and it is hypothesized that there is a need to assess language learners' learning outcomes based on the principles of the Common European Framework in an attempt to have a strong and standardized language teaching and learning environment.

A CEFR-based assessment system can be of use for language teaching departments at universities in Turkey. Throughout this study, it is assumed that the data to be gathered will provide great help to achieve the desired writing assessment objectives at A2 (Waystage) Level.

Scope of the Study

Most language learners in the preparatory language schools of universities are required to reach at B1 level through a one-year English language program to pass their English course and then carry on with their education in their field of study, which makes learning English the first crucial phase of university education. Considering all these, writing skill has an important place in language proficiency since it is a way to explain, educate, inform, entertain one's natural, social behavior, professional and personal point of view or philosophy or expertise to the specific or broad audience by using words and appropriate vocabulary in a systematic method and flow, so assessing writing is essential as it provides diagnostic information about skill strengths and weaknesses (Council of Europe, 2001). Therefore, this study was conducted to determine how effectively the writing tasks of the A2 Level students are assessed by using the CEFR descriptors and the focus group of this study is the university students, those at Karabuk University.

Limitations

This study has been carried out on those prep class students studying English at the SFL Department of Karabuk University. There seems to be a lack of separate writing assessment rubrics for each level, so after the writing tasks are presented to the focus group, the task papers will be assessed by the focus group instructors offering writing course via the writing assessment rubric they have been using. Then, the rubric in question will be analysed to find out if it is a CEFR based one or not, and the need for the standardized assessment system will be tried to be met by this way.

In addition to this, the possibility that the students may not be honest while answering the questions in "can do" statements seems to be another limitation of this study.

Definitions

The definitions of the following terms by The Impel Glossary are as follows (The Impel Glossary, 2007):

Basic User: A term used to define beginner and lower level learners at the CEFR descriptors.

Can-Do Statements: The explanations that are used to inform the learners about their performance in a certain skill.

Common European Framework of Reference: A 260-page long common basis to have a standardized quality and transparency in language learning and teaching area. It is a guiding source to provide a solid basis for the design of syllabuses, curriculum, textbooks and examinations. It also states six levels of proficiency, which are adopted in Europe.

Common Reference Levels: These six broad levels defined in the framework help get a standardized system for describing the language skills learners are expected to have at each level.

A1 (Breakthrough): The lowest level which corresponds to foundation.

A2 (Waystage): The second level which corresponds to basic user.

B1 (Threshold): The third level which corresponds to independent user.

B2 (Vantage): The fourth level which corresponds to the limited operational level.

C1 (Effective Operational Level): The fifth level which corresponds to adequate operational proficiency.

C2 (Mastery): The sixth and the last level which corresponds to comprehensive operational proficiency.

Council of Europe: A political and cultural organisation having 47 member countries. The head office is in Strasbourg, and it is run by the Committee of Ministers.

Descriptor: A definitive statement in which the features of production by the learners are defined.

Conclusion

"Evaluation takes not only teachers, but also learners to create a learner-centered curriculum" (Nunan, 1988, p. 116). What is aimed in this study is to have both learners and instructors have a chance to benefit from the CEFR descriptors to monitor their own learning and teaching, and thus improve learners' language skills by the given feedback. The study is to focus on the following points based on the Common European Framework.

- ✓ Can-Do statements to determine focus group students' level,
- ✓ A writing assessment rubric to evaluate the writing tasks of the focus group, and compare with the CEFR,
- ✓ An interview to see what the instructors' perceptions on the CEFR are,
- ✓ Conformity of the gradings by the instructors offering writing course and the standardisation team.

Chapter 2
Literature Review
Introduction

It took the Council of Europe 4 decades to bring the Common European Framework of Reference out to identify the different language learning levels and skills associated with each proficiency level. The project has been approved and supported passionately by some authorities while criticized grievously by others (Little, 2005). Yet above all, what made it a must was the fact that the framework became the most-awaited standardised evaluation tool for language proficiency levels by promoting plurilingualism and developing educational exchanges (Council of Europe, 2001). It takes action like a common basis for every language educator for the elaboration of examinations, curriculum guidelines, textbooks, etc. all across Europe and it is highly recommended by the CoE to reconstruct language learning process with regard to textbooks, curriculum, examinations, system, etc. (Ahuoglu, 2007). Wernicke (2014) also emphasized the CEFR's importance on having learner-based objectives, self-assessment, autonomous learning, task based teaching, communicative activities and "Can Do" statements.

As North (2014) put it, the most crucial function of the CEFR is its inspiring and motivating reflection on latest practice, and hereby encouraging improvement in both language teaching/learning and testing. The CEFR should not be considered as a harmonisation project, but a reform and innovation movement for networking as it is stated in the CEFR.

In the assessment part, the CEFR provides us with the examples of contexts and purposes, which enables us to categorize learners' needs in given situations. In that descriptive scheme, language use is divided into four domains (Council of Europe, 2001, p. 45): personal, public, occupational and educational. Considering the purpose of this study, educational domain is targeted in order to assess classroom written products for diagnosing language ability. With this thought in mind, there is a need of on-going standard setting process. To have a better understanding on the aforementioned objectives, the assessment of writing skill, standardisation studies on it, the history of the CEFR and the studies carried out on the CEFR are dealt with in the following.

The Common European Framework of Reference

        Introduction. Europe, with the integration of European countries, has come to be a multicultural and multilingual continent to develop a perception of political unity, and along with the development of communication technologies, it has become possible to reduce distances. Change and reconstruction should be the purpose of educational policies, which necessitates European society to be open, plurilingual and cooperative to meet the needs of all citizens (Moreno, 2003). For this reason, the Common European Framework of Reference was published by the Council of Europe in 1998. As Little (2006) put it; this descriptive scheme would be of great help in determining L2 learners' needs, providing guidance on the development of L2 activities and materials and identifying L2 learning goals.

        Historical Background of the CEFR. The Council of Europe has worked on the issues related to language teaching and language learning for four decades, so the CEFR did not come out fully formed. The symposium about languages in adult education being held in Switzerland in 1971 was the starting point of practicing the CEFR in Europe (Little, 2006). After the symposium, the Council of Europe took action in developing syllabus specifications for different language learning levels. First, the Threshold Level was proposed, which caused an enormous impact all throughout Europe, then Waystage and Vantage followed (Council of Europe, 2002).

        After the symposium in 1971, the Council of Europe went about creating a coherent framework and putting different levels and their guidelines together into it. The Framework came in view with the great effort put into practise by Dr. Brian North, Joseph Sheils, Daniel Coste and Dr. John Trim (Council of Europe, 2002). 292 foreign language teachers from Switzerland worked collaboratively and a pool of descriptors was formed, and most of these descriptors were taken from existing scales to assess levels of language proficiency (North, 1993).

        Two drafts were published; the first one in 1996, and the second one in 1998. Wide-ranging feedback from users and discussions all led to the publication of the Framework's current version in 2001 by Cambridge University Press with the European Year of Languages (Council of Europe, 2002). In 2006, on its

website, the Council of Europe announced 21 languages translations of the Framework other than English, French and German (Little, 2006).

The Framework was tested during 1997 and 2000. Brief guides were published to publicize it, then most of those guides were gathered into two volumes. The first one (Council of Europe, 2002) is a general introduction to the Framework for learners, teachers, teacher trainers, language curriculum developers, textbook and language learning material designers while the second one (Council of Europe, 2002) contains test development and language assessment (Little, 2006).

In 2001, in Cracow Poland, 47 ministers of education from different countries in Europe attended an official meeting, at the end of which the CEFR was embraced as the guiding source for all ministers belonging to member countries. It was declared in that meeting that portfolio-based assessment and standardisation in language teaching and learning through the principles of the CEFR should be the focus point in language teaching and learning studies, and that declaration triggered large-scale studies all around Europe, which is considered to be a corner stone as it was the first structured and systematic attempt to enable language teaching and learning activities with a standardized way regarding proficiency. (Little, 2006).

The Need for the CEFR. After the European Union was founded, the interaction between the citizens from different countries, languages and cultures increased accordingly. As the citizen mobility between different countries increased, the Council of Europe specified the need for a new policy on education. On the other hand, xenophobia 'ultra-nationalist' backlashes could be a big problem and bring European mobility to a halt (Council of Europe, 2001, p. 4).

With these thoughts in mind, a series of measures as in the following were set out by the Council of Europe:

- To qualify all European members with international mobility and co-operation not just in culture and education, but in industry and trade, as well.
- To create international communication atmosphere and foster respect for cultural diversities and identities.

- To develop mutual knowledge on regional and national languages, and with this, to expand the diversity and richness of European cultures.
- To prioritize cultural and linguistic studies so that all Europeans communicate with each other and the need for multicultural and multilingual Europe is met (Council of Europe, 2001, p. 3).

In view of these objectives, the Council of Europe remarked the importance of strategies to promote language learning with the aim of advocating plurilingualism, information and communication technologies (Council of Europe, 2001).

As stated earlier, the CEFR serves like a common ground for all course/ curriculum designers, teacher trainers, teachers, language teaching centres, educational administrators and testers. The elaboration of language curriculum, exams, textbooks, syllabuses, etc. all across Europe and the improvement of the quality in syllabuses and the transparency of the courses, which will lead to further progress in international co-operation, are the main purposes of the Framework (Council of Europe, 2001).

Moreno (2003) also emphasizes the importance of the CEFR as it has;
- the related skills and knowledge;
- the communicative strategies and activities;
- the domains and situations of communication;
- the competences needed for communication.

The CEFR is being used widely now, and day by day, its practices are embraced by many countries. Its positive and systematizing effect on educational systems is undeniable. It is believed that the Framework will guide us more with time, and give us the chance to create a standard, CEFR-based basis in every educational corner.

CEFR and Writing. 30 years of experience lies behind the CEFR to develop syllabuses, curricula, teaching materials for different proficiency levels. However, the question was whether widely used practical perspectives in the CEFR were supported with empirical studies or not. Many colleges and ministries of education regard the CEFR as necessary to have their exams standardised. Many of these

institutions claim that their examinations are designed by taking the language proficiency levels of the CEFR. Yet still, few of these proved the case with empirical evidence (Alderson, 2007).

One of the empirical studies carried out on the CEFR was performed by Riikka Alanen, Ari Huhta and Mirja Tarnanen at the University of Jyväskylä. In their research, they worked on the Project Cefling (The linguistic basis of the Common European Framework levels: Combining second language acquisition and language testing research) which was set up with the purpose of integrating language testing based on the Framework and second language acquisition across Europe and working on the linguistic features of the levels stated in the CEFR scales. Also, it is important to note that this project focuses on only writing. They wanted to describe language features that L2 learners make use of at various proficiency levels. With this thought in mind, a set of four L2 writing tasks from young and adult L2 learners in Finnish and English were collected. Two rubrics - the National Certificates (adult learners) and the CEFR - were used to rate learner performances. The CEFR scale consists of six CEFR scales for writing, the National Certificates, on the oher hand, is adapted from the CEFR scale. The only difference between the two scales was found that the CEFR has genre-specific level descriptors to be used for different text types while the National Certificates does not. With the data gathered, a learner corpus to analyze the linguistic features of the CEFR was built (Alanen, Huhta & Tarnanen, 2010). This study is also important in terms of comparing two scales.

In the other phase of the study, 3427 L2 English and Finnish scripts were gathered from different schools, and 1789 of them were selected to be assessed by a group of 9 English and 11 Finnish trained raters. Each script was rated by three or four raters. The raters were requested to state the level they assumed each task would be the most suitable for, and then the second most suitable for. Direct observation to determine rater agreement was the way in the first Cefling study to analyze the relationship between proficiency (CEFR) levels and linguistic features. To carry out linguistic analyses in Cefling, the writing samples that the raters were agreed on were chosen, which means two out of three raters in Finnish and three out of four in English agreed the scripts belong to the same proficiency level. An extra criterion was also made use of so as not to have the

remaining raters diverge from others more than one CEFR proficiency level up or down. Unless these criteria were fulfilled, the sample script was not included in the data, so 63% of the rated writing samples in English and 92% in Finnish were included in the data set (Alanen, Huhta & Tarnanen, 2010). In this study, it is important to note that the scales used by the raters has a crucial place in having a standardised assessment process.

In another study by Forsberg & Bartning (2010), It was aimed to look for the linguistic proficiency features in written L2 French such as discourse organization, formulaic sequences and morpho-syntax. Accordingly, the study shows linguistic profiles of written productions which were rated as belonging to the same CEFR levels. The written data were gathered from 42 university students of L2 French in Sweden during 2007-2008. The students were placed on a CEFR level based on their production of written summaries and argumentative texts and the grading was carried out in accordance with CEFR criteria, narrow linguistic analysis and raters' judgements (Forsberg & Bartning, 2010). The CEFR raters in that study used both Finnish National Certificates, which is based on the CEFR and the more language-oriented criteria suggested in the manual *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Council of Europe, 2009)*.*

The students, most of whom were at B1 level according to the DIALANG test, were requested to perform three tasks in 2-2.5 hours of lesson in a computer room. No aids such as books, dictionaries or grammar check were allowed. The written tasks were modelled on the ones in the Cefling project. As for the rating, professional CEFR raters, who took communicative function and linguistic form into consideration, carried out the process to see if the students performed at the level they had been tested for. All 83 productions were rated by one main French rater, and some productions were rated by a second rater to ascertain the decisions of the main rater (Forsberg & Bartning, 2010).

As a result, 83 productions by 42 writers were grouped. The results reveal that morpho-syntactic measures bring in significant differences while lexical formulaic sequences were found to increase at higher CEFR levels (Forsberg & Bartning, 2010). The rating was carried out in accordance with the two criteria

stated above, and it can be said that the difference between raters is not significant.

At the University of Jyvaskyla in Sweden, Palviainen, Kalaja and Mantyla (2012) studied L2 writing development and analysed fluency in respect to proficiency. The university students who were taking English or Swedish as their major subject were required to write one narrative and one argumentative text. Of the texts, 41 were written in Swedish and 62 in English. A keystroke-logging programme recorded the writing process with all revisions and pauses. The texts were assessed by trained raters by using a standard six-point scale based on the CEFR (Palviainen, Kajala & Mantyla, 2012).The scale was a compilation of criteria such as Overall written production; Creative writing; Written interaction, Coherence & cohesion and Thematic development presented in the CEFR (Council of Europe, 2001) on writing different types of texts.

To ensure reliability, at least two out of three raters of the texts in Swedish, and three out of four raters of the texts in English had to reach an agreement on the CEFR proficiency level. Besides, the rater deviating from the others could do so by only one level. 13 texts (out of 103) were rated at B1 level, 15 at C2 level, 31 at B2 level, and 44 at C1 level. And it was found that the relationship between fluency and L2 writing development is not inevitably a linear one (Palviainen, Kajala & Mantyla, 2012).

A research to investigate criterial discourse features in L2 writing by analysing rated learner essays across CEFR B1, B2 and C1 proficiency levels from Chinese learners was carried out by Chen and Baker (2016). Experienced raters rated the students' expository and argumentative essays and then those essays were put through post-rating statistical analysis. Out of the essays, three subcorpora representing B1, B2 and C1 levels of the CEFR were chosen to be investigated (Chen & Baker, 2016).

The standardizing procedure of the judgements that was used in that study was formulated from the manual for *Relating Language Examinations to the CEFR* (Council of Europe, 2003). A rating scale from the manual, involving three analytical criteria and overall descriptors was made use of in holistic scoring. To put it into phases: in the first one, there was a familiarization training on the CEFR,

then appropriate samples from essays, which were considered as representative ones of the CEFR levels, were selected. Three experienced raters were trained on the standardisation of the essays. A post-standardisation marking test which consisted of appointed essays in a CEFR level was passed by those three raters, two of the raters solitarily marked the same 1,009 essays. Essays which were assigned different ratings were given to the third rater to be marked again. By this way, essays in question got two or three ratings. For inter-rater reliability, to decide if each of them would be included in the subcorpora which is CEFR-aligned, a statistical analysis was performed (Chen & Baker, 2016).

As for results, inter-rater reliability between two raters was found 0.844 whereas it was lower at 0.766 when the ratings of the last rater were included, which was because the third rater only rated the ones that got different ratings from the other raters. In the event of disagreement between the raters, the essays the fit value of which was higher than 1.3 were excluded. Three learner subcorpora embodying B1, B2 and C1 CEFR levels were established after the robust rating procedure and a 202,154 word corpus with the total number of 585 essays was formed. It has been found out in the study that more features with conversation are shared in the writing of lower level learners while academic prose is observed in the discourse of competent writing (Chen & Baker, 2016). In this study, the raters had to reach an agreement on the grades, but what also makes this study important is the training of the raters. This is an indication of the importance of it.

The Common Reference Levels. The CEFR gains acceptance as a descriptive scheme the purpose of which is to analyse learners' needs in learning a language, to draw their learning goals up, to lead the way to develop activities and learning materials involved, and to bring forth orientation to assess learning outcomes (Little, 2006, p.167). The Common Reference Levels help serve this purpose. Glover, Mirici and Aksu (2005) compared Common Reference Levels with the traditional terms; beginner, pre-intermediate, upper-intermediate, etc., and found Common Reference Levels more 'user-friendly' as they make it easier to assess the learners in each level concerning all skills, plus spoken interaction. (Glover, Mirici & Aksu, 2005). With the aim of enhancing the CEFR's usability, three main user levels have been built:

- *The basic user* – the most basic expressions in language; interlocutor's support is needed – the interlocutor's willingness to adjust to the attained level is necessary;

- *The independent user* – the interlocutor can easily deal with daily speech, interact with others with less effort;

- *The proficient user* – the interlocutor hardly has any trouble while using the target language (Martyniuk, 2006, pp. 8-9).

Considering the learning outcomes of the European language learners, six broad levels are outlined as a framework by the Council of Europe (Council of Europe, 2001, p. 23);

- **Breakthrough**, fitting in with the 'Formulaic Proficiency' that Wilkins put forward in his proposal in 1978 and the 'Introductory' by Trim in the same proposal.

- **Waystage,** mirroring the content specification by the Council of Europe.

- **Threshold,** mirroring the content specification by the Council of Europe.

- **Vantage,** mirroring the third content specification by the Council of Europe. Wilkins suggested this level as 'Limited Operational Proficiency, and Trim stated that it is 'sufficient reply to the conditions that learners face with'.

- **Effective Operational Proficiency,** Trim called it as 'Effective Proficiency' while Wilkins went for 'Adequate Operational Proficiency' that embodies advance level competence appropriate for more complicated study and work tasks.

- **Mastery,** Trim called it 'Comprehensive Mastery' and Wilkins 'Comprehensive Operationa Proficiency'. It fits in with the primary exam objective in the program that ALTE endorsed. It can be expanded to involve more improved intercultural competence which is not easy to be managed by most language experts.

"The Common reference levels of the Council of Europe provide a common standard against which the assessment of modern language attainment in different educational sectors, target languages, linguistic regions and states can be referenced" (North, 1999: 25 cited in Shneider & Lenz, 2006: 41). And this accepted standard is elucidated by;

- The Global Scale (CEFR: Table 1)
- The Self-Assessment Grid (CEFR: Table 2)

In the CEFR, with the purpose of making the system easier for curriculum planners, teachers, non-experts, etc., a simple, easy-to-use global scale was introduced. What makes it important is that it was prepared in six levels (as mentioned above), and it also shows what kind of language skills at what extent should be required in each level (Council of Europe, 2001, p. 24).

Table 1

Common Reference Levels: Global Scale (Taken from Council of Europe, 2001)

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating other shades of meaning even in more complex situations.
	C1	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.
Independent User	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.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.

Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	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.

Since there is a wide consensus over the issue of the levels' nature and number suitable for the language learning organization, we have six levels (Council of Europe, 2001, p. 22). But still, the Swiss Research Project came up with a scale of nine almost equally sized as A1, A2, A2+, B1, B1+, C1 and C2. The illustrative scales show the plus) levels labelled as A2, B1, and B2 in the following (Little, 2006, p.168).
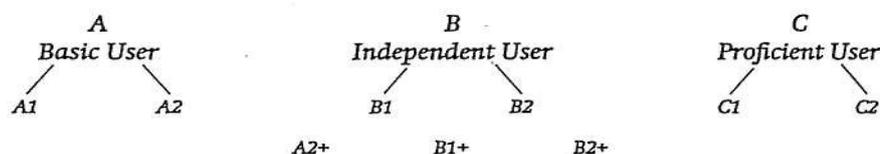
Figure 1. The plus levels (council of europe, 2001).

Other than the global scale, there are various kinds of illustrative descriptors and scales. One of the most crucial scales happens to be the self-assessment grid, that is to say, can-do check list which was brought out by the Association of Language Testing in Europe (ALTE). The grid is divided into three categories which are understanding, speaking, writing, and in four sub-categories as spoken production, spoken interaction, reading, listening and writing language skills in six levels: A1, A2, B1, B2, C1, C2 in the CEFR (Council of Europe, 2001, p. 25). In this study, we deal with A2 Level, so we see the levels A1, A2, and B1 in the self-assessment grid below. The Council of Europe's website can be visited to check the self-assessment grid and the global scale.

Table 2

Common Reference Levels: Self-Assessment Grid (Taken from Council of Europe, 2001)

	A1	A2	B1
Listening	I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to the areas of most immediate personal relevance. I can catch the main point in clear, simple messages and announcements.	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programs on current affairs or topics of personal or professional interest when he delivery is relatively slow and clear.
Reading	I can understand familiar names, words and very simple sentences (e.g. on notices, posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable in simple everyday material such as advertisements, procpectuses, menus and timetables and I can understand short simple personal letters.	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.
Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my eductional background and my present or most recent job.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.

	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	I can write short, simple notes and messages relating to matters in areas of immediate need. I can write a very simple personal letter, for example thanking sb for something.	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.
Writing			

In brief, the Common Reference Levels can be presented in a couple of ways. Yet, they are supposed to have three features to provide the development and future planning: transparency, coherence and being a tool (North, 1994). These descriptors have the purpose of providing a sound basis evaluation instrument in order to improve themselves, evaluate their progress and help teachers plan the language learning process.

Content Coherence in Common Reference Levels. Learners need to pursue and implement the reference levels in the framework so as to make good use of their language learning studies (Little, 2005), and as put by Glover, Mirici and Aksu (2005), since the Reference Levels are used by various organisations and institutions as a useful tool not only for formal but also for informal assessment, this makes the use of the Reference Levels a must (Glover, Mirici & Aksu, 2005). In terms of countries, it is needless to say that they all have different education systems in their appearance. Yet, they should all possess the six levels and five skills determined in the framework. (Little, 2005).

The content of each and every level in the framework is analyzed as the following (Council of Europe, 2001, pp. 33-36)

Level A1 (Breakthrough) – the lowest level of language use. In contrast to using lexically organized phrases, the learner can perform simple tasks in language. In this level, learners are supposed to:

- interact in a very simple way;

- ask and answer very simple questions about themselves such as where they live, things they have and people they know;
- start and respond to simple sentences in case of a sudden need or on very familiar topics.

Level A2 (Waystage) – the level referred to transactional specifications and social functions in 'The Threshold Level' for those people who live in different countries. Learners are supposed to:

- use very simple everyday sentence forms for greeting and address;
- ask how people are, greet them, and react to news;
- deal with short social exchanges; ask and answer questions about what they do in their free time and at work;
- respond to invitations and make one if needed;
- discuss where to go, what to do and make arrangements;
- make and accept offers;
- make very simple transactions in post offices, shops or banks;
- get simple information on travelling;
- use public transport such as taxi, train or bus; ask for information, ask and give directions;
- provide everyday goods and ask for services (Council of Europe, 2001).

Level B1 (Threshold Level) – specification especially for a visitor to a foreign country, and it is categorized by two features. The first one is having the ability to keep interaction and make what the user wants to do in different contexts clear. For example:

- generally follow points of discussion around him/her, speech is provided obviously in a standard dialect;
- give personal opinions in a discussion with friends; express and explain the main points he/she wants to perform comprehensibly;
- make use of a wide range of very simple language to express much of what he or she wants to;
- maintain a discussion or conversation, yet may sometimes be hard to follow when striving to say exactly what he/she likes to;

- keep going comprehensibly, although pausing for lexical and grammatical planning and repair is very evident, especially in longer stretches of production (Council of Europe, 2001).

At this level, the learners have the abillity to keep up a conversation even though they might pause to plan the grammatical and lexical structures. The second feature happens to be the ability the user possesses to deal with the problems in daily life easily. To illustrate:

- handle less routine circumstances on public transport;
- handle situations likely to happen while making travel arangements via an agent or while actually travelling;
- enter into conversations on familiar topics without any preparation;
- complain;
- perform some initiatives in a consultation or interview, yet is very dependent on the interviewer in the interaction;
- ask someone to elaborate or clarify what they just said (Council of Europe, 2001).

Level B2 (Vantage Level) – a new level far above B1. That implies that the learner has been in a slow progress, but it is steady, leading him/her across the intermediate level. The learner confronts a discourse focusing on argument. Learners at this level are good at using a number of interconnected devices to link sentences and make the relationship between ideas clear. For example:

- sustain and account for his/her opinions in a discussion by providing explanations, comments and arguments;
- explain and exploit a viewpoint about a topical issue stating the advantages and disadvantages;
- develop an argument by giving reasons to support or be against a view;
- explain a problem deeply and make it obvious that her/his counterpart in a negotiation has to make a concession;
- speculate about consequences, causes and hypothetical situations;

- take part in an informal discussion in familiar contexts, putting point of view clearly, commenting, evaluating alternative proposals and responding to hypotheses (Council of Europe, 2001).

With time, the learner sees the improvement she/he has and encounters two new focus areas. The first one is about social discourse, some examples of which are:

- converse naturally, effectively and fluently;
- understand what is said to her/him in detail in the spoken language even if in a noisy environment;
- start conversation, take her/his turn when suitable and end it when she/he needs to although she/he might not always do that elegantly;
- use stock phrases to gain time and by this way keep the turn while formulating what to say;
- interact with a certain degree of spontaneity and fluency making regular interaction with the native speakers quite possible without causing a problem on either party;
- have relationships with the native speakers without unintentionally or not amusing or irritating them or asking them to behave other than they would with another native speaker (Council of Europe, 2001).

The next focus is having language awareness with a totally new degree. It can be clarified with the objectives stated below:

- correct mistakes if those mistakes have led to misunderstandings;
- make a note about "favourite mistakes" and monitor speech for it/them;
- generally correct errors and slips if she/he has become conscious of them;
- plan what is to be said by considering the effect of it on the recipient(s) (Council of Europe, 2001).

Level C1 (Effective Operational Proficiency) – good access to a wide range of language which comes with fluent and spontaneous communication. At this level, the user:

- can express and explain her/himself in a fluent and spontaneous way, almost with no effort;

- has a command of a large lexical repertoire that allows gaps to be overcome with circumlocutions;

- there is little obvious searching for avoidance strategies or expressions; only a conceptually hard object can hinder a smooth, natural flow of language (Council of Europe, 2001).

Level C2 (Master) – a degree of accuracy, and no difficulty in the language. The label 'Mastery' does not mean that the user is a native speaker or has near native-speaker competence. The speech of the learners who are highly successful ones is characterized here. Descriptors for this level can be:

- come up with finer shades of meaning by using reasonable accuracy and a wide range of modification devices;

- has a command of colloquialisms and idiomatic expressions with the awareness of connotative level of meaning;

- restructure and backtrack around a difficulty so smoothly that the interlocutor becomes hardly aware of it (Council of Europe, 2001).

In this part, the importance of reference levels is emphasized. As seen above, the progress from the lowest level to the highest one is so smooth. With the details provided in the Framework about each level and what learners are supposed to accomplish at them, it is easier to follow the path as well as to give feedback to the learners. By this way, different institutions have common, standardised education systems.

Can-Do Statements. Can-do statements are considered as central elements of the CEFR. As Heyworth (2005) puts it: "Can-do statements are referred to as an action-centered view of language learning and use" (Heyworth, 2005, p. 12). Can-do statements are here for expressing what learners can do for six levels and the following six language activities that included in the descriptive scheme, and 'can-do' descriptors are used to specify the Common Reference Levels as in the following (Council of Europe, 2001, pp. 25-26):

- Listening;

- Reading;
- Spoken Interaction;
- Spoken Production;
- Written Interaction;
- Written Production.

Combining these language activities and relating them to the six levels bring a self-assessment grid up (see table 2) with descriptors of learning outcomes. For each reference level, it is possible to see a number of can-do satements worded in a positive way. By this way, even learners in the lowest level can perform certain tasks. The competences stated in the statements identify what non-native speakers can do. To exemplify, the general descriptor for writing on *Waystage Level* (or level A2) is formulated as follows (Council of Europe, 2001, p. 26):

"I can write short, simple notes and messages relating to matters in areas of immediate need. I can write a very simple personal letter, for example thanking someone for something."

Another example below is the general descriptor used for spoken production on *Threshold Level* (or level B1):

"I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions."

With Can-Do Statements, learners get the feeling that they achieve something in the target language step by step. Each level informs learners and teachers on what learners are supposed to achieve. As these statements are also a crucial way of feedback, learners have the chance for the immediate compensation for their lack of proficiency in any level.

Testing and Assessment Practices According to the CEFR Principles

Assessment, as put by Piccardo et al. (2009), is a part of language learning and teaching, not just a final act in the process nor a judgment on an activity achieved (Piccardo et al. 2009, p. 41). It is a way to determine the proficiency of learners in target languages. It is fair to say that all language tests serve as a way

of assessment, yet many forms of assessment such as checklists, teacher observation would not be considered as tests. On the other hand, there are many different kinds of assessment. It would be a terrible mistake to think that one type is superior to another one (Council of Europe, 2001, pp. 177-178).

It has been pointed out that CEFR's contribution to assessment is huge (Coste, 2007; Fulcher, 2008; Little, 2007). It is for certain that the publication of it and successive pilot version of the manuel which is a guide to establish a connection between the CEFR and examinations have become a great help for language testers (Council of Europe, 2003). Also, governments and test users have started to make use of it in their education systems. This influence keeps increasing, and with regard to the use of the CEFR, there is a recent call by the Council of Ministers (Council of Europe, 2008, p. 12) requesting the attendance of countries to:

> "ensure that all tests, examinations and assessment procedures leading to officially recognised language qualifications take full account of the relevant aspects of language use and language competences as set out in the CEFR, that they are conducted in accordance with internationally recognised principles of good practice and quality management, and that the procedures to relate these tests and examinations to the common reference levels (A1-C2) of the CEFR are carried out in a reliable and transparent manner".

In assessment, three concepts are considered pivotal, and these are: *validity, reliability and feasibility*, so it is certainly necessary to have a look at them to see their relevance to the CEFR (Council of Europe, 2001, p. 177).

- *Validity:* The Framework is concerned with this concept. A test or assessment process can be considered to have validity in terms of the degree that it shows what is intended to be assessed is actually assessed, which is the necessary information to see the proficiency of the language users.
- *Reliability:*This concept is a technical one. It indicates the extent to which a candidate's same rank order is duplicated in two different administrations of the same assessment.

- *Feasibility:* This concept is about performance testing. Assessors feel pressure while carrying our an assessment procedure as they just see a limited part of the performance, and there are just a small number of categories and types as criteria. At this point, the purpose of the Framework is to provide reference rather than a practical assessment tool.

The Framework as an Assessment Resource. 292 Swiss language teachers who were working in different fields of language education such as vocational training, adult and secondary education initiated the use of the Common European Framework of Reference. The existing scales that had been used earlier to describe the levels of language proficiency were used by them as the starting point of the framework (North, 1995). On the selection, formulation and categorisation of these descriptors, teachers provided great help thanks to their experiences, after which the descriptors were scaled on the six levels of language proficiency (North, 1995), and since then, as Janssen-van Dieten states, the framework's purpose has been to serve as a "tool for assessment, achieve more coherence and harmony in the field of languages and comparability of language qualifications within the European Community" (Janssen-van Dieten, 2003, p. 143).

In the CEFR, the scales and the descriptors are fundamental parts of the assessment issue. The former serves as an informant to develop rating scales which are to be used in the assessment of a learning objective. The latter, on the other hand, is there to provide guidance in the formulation of criteria. The objective may be a specific alignment of skills, activities or competences as well as a wide level of language proficiency in general (Council of Europe, 2001, p. 179).

Scales are considered truly efficient when they define what language learners can do and also how they are supposed to do it. While using scales, functional distinctions between them should be kept in mind. The CEFR identifies three types of scales and their functions (Piccardo et al. 2009, p. 52):

a) user-oriented scales, describing what language learners can do and report about their own typical behaviours;
b) assessor-oriented scales, pointing the importance of the quality of learners' performances and functional for guiding assessment;

c) construct-oriented scales, guiding the construction and formal assessments contents such as tests focusing on what language learners can do.

As for descriptors, it will be necessary to make a distinction between two types of it (Council of Europe, 2001, p. 179):

Descriptors of Communicative Activities. These kinds of descriptors are useful for teacher or self-assessment as they have real-life tasks. Since both teacher assessment and self-assessment have learners and teachers concentrate on an action-oriented approach, they are found effective. Yet still, should one be interested in documenting results concerning a proficiency level, descriptors of communicative activities are not recommended to be used in the criteria to grade a performance in a speaking or wrriting test. The reason for this is that, in order to document proficiency, the assessment is supposed to assess generalisable competences based on the performance, not to be concerned with the performance itself (Council of Europe, 2001, p. 180).

These descriptors are in three different ways with regard to achieve the objectives (Council of Europe, 2001, p. 180):

1. **Construction:** They are used in the definition of a specification while designing assessment tasks.
2. **Reporting:** They are also useful while reporting results. Employers, as the users in the educational system, are highly motivated to see the overall outcomes.
3. **Self or teacher assessment:** Last, they can be used for self and teacher assessment in different ways as in the following:
   - *Checklists:* Checklists are useful for continuous assessment or summative assessment carried out at the end of courses. The descriptors' content can be "exploded". To illustrate, "Can ask for and provide personal information" might be exploded into the implicit constituent parts "I can introduce myself; I can say where I live; I can say my address in French; I can say how old I am, etc. and I can ask someone what their name is; I can ask someone where they live; I can ask someone how old they are, etc."

- *Grids:* Grids are used for continuous or summative assessment while grading a profile with the categories defined for each level in a grid (Council of Europe, 2001, p. 180).

Descriptors of Aspects of Proficiency Related to Particular Competences.
Descriptors of aspects of proficiency are in two different ways with regard to achieve objectives.

1. *Self or teacher assessment:* The descriptors are statements that are in a positive and independent way and can be involved in checklists for self and teacher assessment. Yet, most of the scales have a drawback that the descriptors, especially the ones at lower levels, are worded in a negative way.

2. *Performance assessment:* The descriptors are of great help for the assessors to be involved in the process in order to improve a common frame of reference (Council of Europe, 2001, p. 181).

Self-Assessment. Students' self-reflective and management abilities have an important role in learner autonomy (Little, 2008). In this view, learners monitor a series of acts in language learning, and they choose technigues, define objectives, know how to learn and determine strategies for success (Holec, 1990). Considering this, self-assessment is a crucial component of learner autonomy as it gives learners the chance to monitor their progress and to make choices concerning their learning. Moreover, learners get a realistic idea of their abilities, get self-aware of their language learning process and gain more control over it (Alderson, 2005).

The fact that self-assessment plays a positive role in development and learning of professional competence is the reason why educators and researchers took an interest in self-assessment (Dochy, Segers & Sluijsmans, 1999). This interest is seen to have higher education change goals; having students be knowledgeable of their field of study is not the only focus, but providing them with delegable skills in professional life is, as well (Dochy, Segers & Sluijsmans, 1999). As Stefani (1994) put it, one precious skill that students should possess is their ability to evaluate and assess themselves to use the information they get from it for their future profession. Taras (2001), on the other hand, stated that self-

assessment serves as a critical and effective tool providing learning beyond higher education.

The Common European Framework of Reference (CEFR) gathers tasks, language learning objectives, assessments and contents together so as to promote teaching and learning. Within the CEFR, self-assessment is carried out through the European Language Portfolio (ELP), which is a tool contributing to learner autonomy through self-reporting, self-assessment and goal-setting (Little, 2005).

The Common Reference Levels (CRLs), on the other hand, are the tools used in the CEFR for self-assessment (Council of Europe, 2001), which involve a self-assessment grid describing learners' performance at six levels (see Table 2). Learners are presented with 'can-do statements' in five language skills areas at six levels (reading, listening, spoken production, spoken interaction and writing). Those statements involve positive terms in order to encourage learners. To illustrate, the statement for writing at B1 level is:

> I can write simple, connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.

With these statements, learners reflect on their abilities, appreciate their achievements and set goals for future learning (Council of Europe, 2001).

Three reasons are stated for engaging learners in self-assessment. First was defined by Nunan (1988) as a learner-centred curriculum, a cooperative effort between learners and teachers. It fails if learners are included in making decisions on curriculum content, but not on evaluation process of curriculum outcomes. Second was put forward by Oscarson (1989), and he stated if self-assessment is made a supplementary part of evaluation procedures, assessment will be regarded as a shared responsibility by teachers and learners. Therefore, self-assessment has an important role in directing reflective processes which the development of learner autonomy depends on. Third, in a world that languages are used beyond classrooms, self-assessment enables learners to use target language for further explicit language learning (Oscarson, 1989).

Aspects of Self-assessment. While considering assessment, two key aspects should be taken into account: understanding how assessments are used in learning, and how they are used for measurement (Butler & Lee, 2010). Measurement aspect is about measuring learners' degree of their mastery of skills and understanding, and the results are used for summative objectives. The learning aspect, on the other hand, aims to advance students' learning. Students are provided with feedback on their self-assessment process, which helps them become more proficient (Boud, 1995).

Self-Assessment as a Measurement Tool in Language Education. The validity of self-assessment, as a measurement tool, is considered to be under threat because of its inherent subjectivity. That's why, researchers, who examine self-assessment's measurement aspect, has gotten an interest in examining the validity of it. These studies have generally investigated the correlations between the scores that are attained through external measurements and self-assessment scores. Results have shown a few factors responsible for the variability, which are: (1) students' characteristics; (2) the ways items and questions are delivered; and (3) the skill or domain being measured (Ross, 1998).

About the first factor, several influential characteristics of the students have been found to affect self-assessment variability. Secondly, the way that items and questions are constructed appers to affect self-assessment variability, as well. Items linked to the task objectives of students have been pointed out to be more accurate. Thirdly, Rose (1998) carried out a meta-analysis of validation on self-assessment, and deduced that receptive skills such as reading and listening are more accurate than productive skills such as writing and speaking while being self-assessed (Oscarson, 1989).

Self-Assessment as a Means of Advancing Learning in Language Education. In foreign and second language education, learning aspect of self-assessment has gained significant attention. This situation has shown a shift from teacher-centered to learner-centered instruction (Boud, 1995).

Students can monitor their own learning progress, assess their own understanding and become aware of the expectations through self-assessment. They can also see how much effort or work is needed to achieve their goals.

Besides, students feel a sense of control over their progress as they both reflect and evaluate their own performances, which lead to motivation (Paris & Paris, 2001).

As stated above, self-assessment can be practical and influential for both teachers and students. It helps students become more autonomous, which can also ease teachers' assessment burden. The aspects of self-assessment should be considered comprehensively and applied accordingly.

Assessing Writing

Writing in a second or a foreign language to state one's opinion on a subject is a crucial part of learning, and doing so with coherence and absolute accuracy is even greater; sometimes even beginner level students do practice in writing in a foreign language reinforcing the language they have learnt while many English native speakers cannot truly master this skill (Celce-Murcia, 2001, p. 205). By writing, it is possible to communicate messages to readers, which is of great importance in the modern world. Therefore, writing needs to be taken into consideration and encouraged during the studies language learners take (Olshtain, 2001, p. 207).

In the changing world that we teach and learn, differences in writing research studies are the results of perspectives and themes. That means there is progress and we need progress. What Anita Poon (2004) stated below is what we have in almost all dominant classrooms:

> "Typically writing is taught based on a prescribed textbook in primary schools. Most teachers simply stick to the textbook and adopt a very traditional method. A typical composition lesson goes as follows: the teacher teaches the class a sample of writing in the unit, which usually consists of several sentences describing a person or an object. Then, with the help of some guiding questions, the teacher asks the class to do parallel writing, which means to write a similar text by changing simply the names, pronouns, numbers or some details of the original text. Finally, the students copy the answers to the guiding questions in their exercise books, and submit their composition". (pp.307)

Given this kind of traditional method while teaching writing in the classroom, there seems to be nothing challenging in it, because the input does not aim at the right level as put forward in input hypothesis by Krashen (Krashen, 1982).

On the other hand, Kaplan's statement as: "cultural thought patterns in inter-cultural education" (Kaplan, 1996, p. 8) indicates that language and writing are cultural facts, and different cultures form different tendencies in writing. At this very point, Connor (2002) puts it as: "the linguistic patterns and rhetorical conventions of the L1 often transfer to writing in EFL context and thus cause interference at the level of syntax, discourse and phonology" (Connor, 2002, p. 494) That's to say, as EFL students live in the culture and experience social context, they mirror their educational system and produce cultural writing features while writing. Yet, this, in a sense, ignores multiple factors contributing to the product of foreign language writing, "such as L1 writing expertise, developmental aspects of foreign language proficiency, and individual writers' agency" specified in their preferences. These views show that features of writing skill can be caused by multiple factors, and not just cultural influence can be held accountable (Ryuko Kubota & Al Lehner, 2004, p. 12).

On the other hand, as the acquisition of a language skill is considered important, it is equally important to test it, and of course writing is not exceptional. The role of writing in foreign and second language education increases, which causes a demand for reliable and valid methods to assess writing ability (Weigle, 2002). But what does 'assessing writing ability' mean? The answer to this question by Hughes (1989) is that the best way to assess writing ability is to have people write. Considering this, two main constituents distinguish in a test of writing: instructions telling students what to write or writing tasks, and a means of assessing samples produced by students. Yet, before making decisions about scoring procedures, a few key questions should be taken into consideration (Weigle, 2002):

- What is it that we are trying to test? In other words, what is the definition of writing ability considering the objectives of the test – is what we want to know if students form accurate sentences, or if they can use writing for a particular cummunicative purpose?

- What is the reason behind our desire to test writing ability? What are we going to do with the results?

- Who are test takers? In order to give them the chance to show us their performance at highest ability through designed tasks, what is needed to be known about them?

- Who will evaluate the papers, what standards, rubrics or criteria will be used? How can we guarantee that the scoring standards applied by the raters are consistent?

- Who is going to use the results / information provided by our exam?

- What are the obstructions (money, time, labor, materials) that restrict the amount of information collected about students' writing ability?

- What should be known about assessment / testing to have a more reliable and valid exam?

Basic Considerations in Assessing Writing. Writing assessment should also be considered from the point of language testing view, which has conventionally taken notice of defining the meaning of language ability in particular, as a fundamental cognitive ability through the traditional four skills (Weigle, 2002). In the following, different test purposes, actual language use and language performance relationship, the concept of performance assessment and lastly, a test usefulness model are presented.

Test Purpose: Making Inferences and Making Decisions. While designing a writing test, the first thing to do should be deciding on our purpose – why do we want to assess writing ability – what are we planning to gain out of this process (Weigle, 2002)? Two key purposes were stated by Bachman and Palmer (1996) for language tests, and writing tests is a part of it. Making inferences about language ability is the main purpose, and making decisions grounded on those inferences is the second main purpose. As it is not possible to observe the language ability of a person directly, their responses to items in the test are evaluated and inferences about their ability are made (Bachman & Palmer, 1996).

To illustrate, three types of inferences on the basis of a language test can be considered: diagnosis, achievement and proficiency. Diagnosis inferences, which refers to the weaknesses and strengths of students, are made use of by

teachers in an attempt to adapt their instruction to the needs of their students. Achievement inferences, which refers to the extent to which students have reached specific goals, are made use of by teachers to make decisions about promotion and grading on the levels. Proficiency inferences, on the other hand, are made use of while making decisions about placement into different levels, selection for a particular job or admission to academic programs (Weigle, 2002).

Language Use and Language Test Performance. As stated above, a language test should make inferences on language ability, so we need to clarify what language ability means. That is to say, how this ability will be adapted to the real world language use and in a language test should be stated clearly. The ability that we wish to test is refered as **construct**, and while developing a test, one of the most crucial concerns is to define construct. The construct of interest is defined by determining the factors real-world language use involves and which of these factors we want to test. To illustrate, suppose a student wants to write an essay and compare and contrast the works of two famous musicians. This writing task would need knowledge about music. In a composition course, students knowledge on music would not be a part of the construct, but if it was the case in a music history course, this would be a part of their construct definition (Weigle, 2002).

In the field, it was agreed that communicative language ability means interaction between strategic competence and language knowledge aspects, as stated by Bachman and Palmer (1996). They (1996) also put forth a more general taxonomy of language ability components, which includes textual knowledge (knowledge of how to form logical texts by putting building pieces of language together), sociolinguistic knowledge (knowledge of using language in various social settings), grammatical knowledge (knowledge of the vital building pieces of language) and functional knowledge (knowledge of how to use language to manage communicative functions) (Weigle, 2002).

Along with strategic competence and language knowledge, Bachman and Palmer (1996) stated that authentic language use in real communicative settings consists of other concerns as well: personality factors, emotional factors, affect factors and topical factors. To illustrate, suppose students are assigned a task

about writing a letter to an editor. This task necessitates more than language knowledge. To start with, students would entail knowledge about the topic (topical knowledge), feel motivated to do the task (affect) and let their characteristics impact on the selection of language and content (Weigle, 2002).

Writing as Performance Assessment. Any assessment procedure including behaviour observation in real world or a simulation of it, which involves assessing the performance of the ability or raters' evaluation of the performance, is described as performance assessment. As it represents behaviours in real life, it differs from paper-and-pencil tests. To put it into other words, in contrast to multiple-choice tests, performance assessment involves actual writing representing a writing performance (Weigle, 2002).

In language testing, McNamara (1996) came up with a distinction between a strong and weak sense of performance assessment. In the strong one, what is important is the success of the task that necessitates language use, not language use itself. To illustrate, if students are assigned to write a letter of complaint, they are successful if the reader is eager to make up to them. In the weak one, on the other hand, it is just the opposite: the focus of the assessment is not on the success of the task completion, but on the language used. That is, the raters are interested in linguistic features while reading rather than being persuaded.

Most language tests, as McNamara (1996) stated, are in between these two processes. Writing tasks at the weak one have limitations to correspond to real-life tasks, and focus on very limited areas of language ability as they are highly controlled in language and content. On the contrary, strong ones give way to factors such as affect and topical knowledge corresponding more to real-life writing tasks (McNamara, 1996).

Test Usefulness. Bachman and Palmer (1996) pointed out that while developing or designing a language test, the most important concern is its usefulness. Six qualities are used to define test usefulness: construct validity, practicality, impact, reliability, interactiveness, authenticity, which are described in terms of their relationship with writing assessment in the following. These features are all important, but it should be kept in mind that it is not possible to maximize

them all. That's why, test developers can work on a suitable balance among the features.

Reliability is an important concern in testing, and for test validity it is a prerequisite. It means the consistency of measurement across various features of a testing condition. If a test is considered reliable, this means students have the same grades from one rater or prompt to the next, and different versions, occasions or raters do not change the result; students are rank-ordered in the same way (Weigle, 2002). There are also a few factors that influence the reliability in a writing test, and a few of these factors, such as writing samples that students are required to provide, the topic and discourse mode of response that is expected, can be related to the task itself while some of them, such as raters' training, rating scale's nature and raters' experience and background, are related to scoring process (Weigle, 2002).

Construct Validity is described as the appropriate and meaningful interpretations made on the test scores' basis. It refers to the determination process on whether a test is literally assessing what is intended to assess. It is crucial to understand what ability is to be assessed and to what extent so as to have decisions based on the results to be fair. Another important thing is the domain of writing which a test is aimed to generalize to. To illustrate, suppose in a business writing test, test tasks consist of memoranda and letters, this means it will not be possible to generalize it to other business writing genres. That's to say, construct validation depends on the definition of interest for a testing context, and is specific to each test (Bachman & Palmer, 1996).

In testing writing, there are at least three ways that construct validity needs to be demonstrated: (1) the components of writing involved in the construct definition have to be taken into consideration in the scoring criteria; (2) while scoring writing samples, the raters have to keep to those criteria; and (3) the type of writing that is needed to be tested has to be elicited in the task (Weigle, 2002).

Authenticity refers that the writing task that is carried out has to be representative of the writing type students will need in the real world, outside the classroom. This is not problematic in some cases. To illustrate, in an EFL class general-purpose English test, writing tasks that students are expected to manage

might be identified easily: a response to a job advertisement or a letter to a tourist agency. Yet, in some other cases it is a bit problematic. It might be difficult to find a suitable writing task representing a target language use with a group of English-speking learners of foreign languages. Test developers are expected to decide that authenticity is less important than the other considerations in this case (Bachman & Palmer, 1996).

Interactiveness is defined as the type, role and extent of students' characteristics in managing a test task. The characteristics for language testing, as stated earlier, are topical knowledge, language knowledge, strategic competence, how emotional students are while responding to the task and affective knowledge. As these are all engaged in language use, interactiveness counts for language testing. Therefore, an assessment task including just language knowledge gives an idea about how much a student knows about the language, not about how well they can use the language (Bachman & Palmer, 1996).

Impact, as the name suggests, refers to the influence tests have on students. The impact of tests on instruction and curricula, referred to as washback, has also received attention in recent years. Washback has the possibility of being positive or negative. If a testing procedure motivates teachers to adopt practices compatible with the best ones in the field, it is a positive washback. Negative washback, on the other hand, encourages teachers for the backlash practices which have nothing to do with the current thinking in the field (Weigle, 2002). To provide positive washback, the purpose of the test should be understood, results should be believable, the test should be grounded on clearly stated goals, tasks should be in accordance with real-world language tasks and self-assessment should be invested in in the assessment process (Bailey, 1996).

Practicality can be described as the availability of the resources for test-based activities in meeting the resources needed to administer the test. Two important reasons exist as limitations for writing assessment: teachers have the desire to collect as many writing samples as possible from the students with the purpose of sampling the domain, yet it is not possible to do in a limited time; and writing tasks are time consuming to be scored, which cause practicality concerns

make scoring procedures impractical. That's why, it is important to have sufficient resources while designing a writing test (Weigle, 2002).

In this part, it was aimed to cater for an introduction to writing assessment by taking the basics of writing assessment into consideration. It shows that a test developer must consider a various of variables before designing a test of writing.

History of Writing Assessment. Writing has always been a powerful tool affecting all societies, yet it took time for us to perceive this power. In the 21st century, all personal and public activities are organized by written documents. In many cases, writing serves as the tool every activity is performed through. And at school, expanding a learner's capacity to write deepens our, teachers', sense of what we try to do every time (Bazerman, 2008). Herein, this part takes the standpoint of writing, how schooling teaches writing and how it has developed.

The direct assessment of writing has been thought as an action and the inevitable result of multiple-choice testing. Yet essay testing - which should be put as assessment through writing, not of writing - has been on the agenda for thousands of years (Hamp-Lyons, 2002). In Europe, written examinations developed much later. In DuBois' report (1970), it is stated that what introduced the written tests to the West was the Jesuit order, publishing an assertion about writing test procedures in the late 1500s. This way of assessment was considered a perfect way to control the teaching processes in Catholic schools. According to Spolsky (1996), this form of control spread all throughout the French education system and then Europe.

University examinations were carried out orally between the times of Middle Ages and the 19th century (Rashdall, 1895), and this system continues today in the doctoral exams. University education in Britain and Europe always had tutor-student dialogue and debate in seminar style. Then, this speaking-based system began to be replaced by written examinations under the guidance of Oxford and Cambridge universities. Written examinations were even begun to be preferred while choosing civil servants as they became a quick and reliable way to assess the intelligence and literacy skills of civil servant (Hamp-Lyons, 2002). Edgeworth (1888) came up with a method in determining the "true judgement" in a written examination text. His method is not very different from the methods used today:

several competent critics from scholars and authors are used, pooled and a mean score is reported. His work pioneered other studies to decrease the intricacy of the assessment of writing to a task to be made objective. Hartog (1910) summed it up in an entry:

> "It can scarcely be doubted that in spite of the powerful objections that have been advanced against [written] examinations, they are, in the view of the majority of English people, an indispensable element in the social organisation of a highly specialised democratic state, which prefers to trust nearly all decisions to committees rather than to individuals. But in view of the extreme importance of the matter, and especially of the evidence that, for some cause or other (which may or may not be the examination system) intellectual interest and initiative seem to call for a searching and impartial inquiry." (pp.49)

At the end of the 19th and the first years of the 20th century, the methodology of writing assessment changed at universities. Written compositions replaced traditional oral exams at universities like Harvard University. Harvard's method included "a short composition, correct in spelling, punctuation, grammar and expression, the subject to be taken from such works of standard authors as shall be announced from time to time" (quoted in Applebee, 1974). Many were against the Harvard approach, one of which was Fred Newton Scott from the University of Michigan. That opposition gave birth to the USA's National Council of Teachers of English in the year of 1911 (Hook, 1979). In that period, the number of the test takers increased, which led institutions to look for a quicker way to assess written texts. That period is characterized by Lunsford (1986) as follows:

> "Where reading, writing and speaking had once been combined in the pursuit of a student's own academic and social goals, writing was now separated from the other communicative arts. The direct result of this separation was a dramatic loss of purpose: writing became not primarily a means of influencing important public affairs but merely a way to demonstrate proficiency. Divorced from its original purpose in rhetorical instruction, writing shifted its focus from discovering and sharing knowledge to being able to produce a "correct" essay on demand; lost the theoretical framework that related language, action and belief; and became increasingly preoccupied with standards of usage, a tendency that grew, by the turn of the century, into a virtual cult of correctness." (p. 6)

The replacement of the liberal humanistic view of writing ability by the behavioural scientific one went further in the USA than in Europe. Educational researchers started to work on what Paterson (1925) called "new-type" exams. Between the years 1920-1960, a new view of literacy consisting discreet skills to be assessed separately was put forward and became prominent in the USA. Yet, there were some concerns regarding this new view as a threat to the "civilising influence" of education (Thomas, 1931). By 1950s and 1960s, the concept of 'objective testing' and writing assessment were focused on in many educational assessment researches. As Yancey (1999) reported, in a survey in 1952 by Sasser, it was found that most of the educational institutions place their students into writing courses making use of standardised tests; and a considerable number of them used those as the final grade from their writing course.

On the other hand, there were concerns and objections in the UK concerning the reductionist approach on writing. Wiseman (1956) exphasized the importance of validity in his work, and elaborated the "backwash effect", which is still important in British educational assessment and called "washback" today. Wiseman with his colleagues came up with the 'Devon' method - multiple marking of written compositions, and since the 1940s, various forms of the Devon method have been used in the UK.

The mid-1960s was announced as the beginning of a new era for writing assessment by Huot (1990). Lots of encouraging researches were carried out to improve direct writing assessment by Diederich, French, and Carlton (1961). In his study, Huot (1990) also stated "direct testing of writing couldn't have its place till the middle of 1970s because of political and economic situations" (pp. 237-238).

Yancey (1999) divided writing assessment into three 'waves': the first one was between 1950-1970, when 'objective' testing was the way writing assessment was carried out through; the second one was between 1970-1986, when the performed practice was holistic scoring of essays; and the last one 1986-present, with portfolio assessment (p.484).

Alternative Assessment. Testing and assessment are substantially different from each other. Testing is standardised and formal whereas assessment is grounded on the information about students' knowledge and what they are able to

do. To put it in a different way, in testing, scoring and administering procedures are given to the students, but in assessment various methods to collect information at different contexts and times exist (Law & Eckes, 1995). According to Dietel, Herman and Knuth (1991), assessment can be described as any method that is used with the purpose of understanding a student's current knowledge better. And Mitchell (1992) defines testing as "single-occasion, unidimensional, timed exercise, usually in multiple-choice or short-answer form."

Assessment has always been a controversial issue. It appears that more modern assessment methods and alternatives have been increasingly used with the traditional ones in recent years. As the name suggests, alternative assessment methods are alternatives to classical ones (Oproescu, 2018). And traditional teaching, assessment theories and applications have been affected by some learning theories such as project-based learning, multiple intelligences and constructivism (Fourie & Niekerk, 2001). One of the reasons for this is because the definition of learning and its meaning have changed (Shepard, 2000). It can be said that this change consists of a very distinctive approach to teaching and learning process and their stages, including a new one (Daghan & Akkoyunlu, 2014).

The term 'alternative assessment' has been entitled by various terms, such as authentic assessment, portfolio assessment, performance assessment, situational assessment…etc. Educators and researchers prefer to use the term authentic assessment, alternative assessment and performance-based assessment interchangeably. The common principle that all these terms share is that they assess students' performances with various real-life tasks while substituting for traditional testing. They are intertwined smoothly into the daily classroom activities within the classroom context (Gill & Lucas, 2013).

In recent decades, the selection of the tools used for assessment has underwent a radical change, and assessment models at schools and all other educational institutions have been changed into alternative ways, as Kalra (2017) stated. Traditional paper-pencil tests and ways to assess students' learning are becoming an old way since they assess students based on a teacher-centered approach, which was, to a great extent, incomprehensible to students. What are

expected from new approaches to assessment are to clarify deeper aspects of learning and give students the chance to self-monitor themselves (Janesick, 2001).

Buhagiar (2007) stated that alternative assessment methods should be preferred instead of traditional assessment in order to provide learners with better learning opportunities, and why these methods have come into use is because they have been considered as a response to the requirements for reformed assessment. Everyone has the ability to learn and every student deserves to be given the best education, so an assessment method based on selection, accountability and certification is unsustainable to continue to be used.

Alternative assessment includes various unconventional assessment methods, such as project-based assignments, portfolio assessment, classroom-based, authentic assessment and informal performance assessment (Gill & Lucas, 2013), and alternative assessment takes on a constructivist learning view suggesting that learners are active in constructing knowledge rather than choosing or selecting (Dogan, 2011), and the main purpose of the alternative assessment is to find out how learners are approaching, handling and finalising real-life tasks in a specific area (Garcia & Pearson, 1994).

Compared with traditional assessment ways, which focus on learners' skills and knowledge, alternative assessment centers upon problem solving, learners' learning strategies, direct and holistic measurements and task completion (Wiggins, 1998). Hence, alternative assessment supports learners' self-discipline and choices by fostering their trust, vision, compassion and spontaneity (Janisch, Liu & Akrofi, 2007). As Balliro (1993) pointed out alternative assessment was broken out "as a result of lacking tools that can show students' real improvement and their strong strides, and the dissatisfaction of implementers about prevalent assessment tests." Furthermore, alternative assessment makes use of activities in order to find out what learners are capable of doing with the skills and knowledge they acquired through learning (Oliver, 2015).

Alternative assessment referred to be qualitative, performance assessment, informal or classroom-based is a system to evaluate student learning other than

formal testing, and it has a few particular characteristics (Janisch, Liu & Akrofi, 2007):

- Alternative assessment is set in the classroom where teachers make choices in the measures that are used.
- In alternative assessment, text, student and context all affect learning outcomes as it is based on a constructivist view of learning.
- Alternative assessment is grounded on the idea that the resulting products are equal to learning processes.
- In alternative assessment, students are asked to produce, create or perform something.
- In alternative assessment, problem solving skills and higher-level thinking are exploited.
- In alternative assessment, it is important to make use of tasks illustrating purposeful instructional activities.
- In alternative assessment, not machines but people do the scoring.
- Alternative assessment requires new assessment and instructional roles for teachers.

As stated earlier, alternative assessments emphasize the significance of examining the processes along with learning products. They give students the chance to explore the possibilities in complex problems and go beyond 'the one right answer'. In the following, the range of alternative assessments are presented. Some of them are declared as new alternatives, yet still they actually are assessment techniques that teachers have coped with for many years. Many teachers use a series of information sources to find out how well students have learned and who is having difficulty. What is new on these assessments is that what was informal and implicit earlier has been turned into formal and explicit. The teachers are also encouraged to state their instructional goals in a clear way, work on their instructional purposes and have alignment between these purposes and their teaching (Herman, 1992). Assessment alternatives are as follows:

Table 3

Alternative Assessments.

Assessing Processes	Assessing Products
• Clinical interviews • Documented observations • Student learning logs and journals • Student self-evaluation (oral or written) • Debriefing interviews about student projects, productss, and demonstrations (student explains what, why, and how, and reflects on possible changes • Behavioral checklists • Student think-alouds in conjunction with standardized or multiple-choice tests	• Essays with prompts and scoring criteria • Projects with rating criteria • Student portfolios with rating criteria • Student demonstrations/ investigations (expository or using the arts) • Paintings, drama, dances, and stories with rating criteria • Attitude inventories, surveys • Standardized or multiple-choice tests, perhaps with section for "explanations"

Teachers should take the following points into consideration so as to increase the effectiveness of alternative assessment (Elliott, 1995):

- Tasks that are connected or aligned to what has been taught should be chosen.
- Before working on the task, students should be provided with the scoring criteria .
- Before students start working on the task, they should be informed about the standards, and provided with a few acceptable performance models.
- Students need to be encouraged to self-assess their performances.
- Students' performances should be interpreted and compared to other students' performances and standards.

As stated earlier, alternative assessment strategies include portfolios, exhibits, computer simulations, open-ended questions, hands-on experiments and demonstrations (Dietel et al. 1991). Portfolios and projects, two common techniques are presented below.

Portfolios. Portfolios include the works of students showing their mastery of task skill (Kulieke et al. 1990). Portfolios are defined by Bailey (1998) as a collection of students' works showing the progress, efforts and achievements of them in a particular area or areas. This collection is supposed to consist of evidence of student's self reflection, his/her participation in choosing contents and the criteria. Portfolios necesitate responsibility of and a lot of input from the student as they are cumulative in their nature. Furthermore, teachers need to commit a great deal of time for them (Bailey, 1998).

Arter (1995) points out the benefits of portfolios as follows:

- Students have the chance to get a broader look at what they know and can do.
- Students ground assessment on an authentic work.
- Students have an alternative to report standardized tests.
- Students have a better way to interact and communicate with parents.

Projects. Individual or group work is possible in projects. So many things can be included in the projects such as learners' prior experiences, real life related concepts and authenticity. What is considered as project? Any type of method showing students' knowledge on a particular subject, such as art work, multimedia presentations, research proposals is considered as project. Students can be given a scenario and required to come up with solutions or strategies since problem-based learning necessitates it for learners to use their problem solving skills. The task can be assigned to individuals or groups. The findings can be presented in many different ways as well such as role-play, presentation or written report (Simonson et al. 2000).

A description of alternative assessment is presented in this part. As stated above, alternative assessments provide students with the opportunity to show their true language acquisition by taking part in projects and situations and performing tasks in the target language. The use of alternative assessments would also shape students' personalities as they get a lot of responsibility of their education.

Approaches. Standardisation for the accuracy in writing scoring is critical. To this end, three approaches exist: consensus estimates, consistency estimates and

measurement estimates, which are all required to provide inter-rater accuracy (Glasswell, Brown & Harland, 2004). According to Stemler (2004), consensus estimates point out the degree that markers give the same scores to, consistency estimates point out the degree "to which the pattern of high and low scores is similar among markers, and measurement estimates indicate the degree to which scores can be attributed to common scoring rather than to error components" (pp.108-110). In commonly used large-scale assessment programs, consensus rates are between 80%-100% while consistency and measurement coefficients range between .70-.80 and .60-.80 (Glasswell, Brown & Harland, 2004).

Consensus estimates are made use of when raters are taught to grade in accordance with the rating scale rubrics believed to embody "… a linear continuum of progress in a construct." Common indices of consensus estimates are per cent exact agreement and per cent adjacent agreement (Glasswell et al. 2004). As Stemler (2004) stated, the adjacent agreement approach gives advantages to the raters to get high scores compared to the exact agreement approach, and having few categories (up to four points) in a rating scale means it is easier to achieve high scores. Yet still, 70 % or more of exact agreement levels are regarded he representative of reliable scoring.

Consistency estimates (i.e. Cronbach's alpha, Pearson) is not about if raters have provided the same scores, but about an obvious pattern in the scores' distribution between raters. If raters give high and low scores in a kind of similar pattern, high coefficient is achieved, but each rater's mean scores can be a lot different because high score for one judge may not be the same point stated on the rating scale. In that case, adjustments for harsh raters can be needed (Glasswell et al. 2004).

Measurement provides the most vigorous indicator of degree of agreement as it estimates "apportion variance in assigned scores to task, rater, error and interaction components" (Shavelson & Webb, 1991). The degree that a point indicates a true score to can be established by such estimates taking rater, error, internal consistency of task and all judges' harshness into consideration. Coefficient values that exceed .80 are considered as the indication of judges' rating a common task (Shavelson & Webb, 1991).

Standardised Writing Assessment. Countries all around the world are increasingly paying more attention to large-scale assessment programs in order to augment accountability of their educational systems, thus standardised writing assessment is becoming a robust issue. Such assessments have lent themselves to improve beyond traditional multiple-choice formats, to alternative assessments which are in accordance with the construct (Shermis, 2014). To illustrate, in States, extended and repeated written performances are a fundamental part of summative and formative assessment of different content areas under the Common Core State Standards. England, on the other hand, has extended written performances as a component of different Key Stage assessments (Shermis, 2014).

The problems with the standardised writing assessment are the problems that arise from the skill itself. Writing involves students' physical skills and complex cognitive processes, which allow them to construct meaning, communicate, express and make connections (Hayes & Berninger, 2014). Writing is an important part of schooling since students might spend half of their class hours attending writing tasks (Brindle & Harris, 2016). The attention span of students to learning writing has been seen to decline in their writing assessment, which concerns education systems all around the world, and this has also encouraged standardised assessments to broaden. As the volume of written performances in standardised assessments has increased, the need to come up with methods so as to better the reliability and validity of writing assessment has also grown (Mackenzie, 2013).

Analytic, rubric-based methods, which involve rating written performances on multiple criteria expected to cover the scope of the writing construct, have dominated standardised writing assessment (Spandel, 2005). In higher education, these rubric-based methods have also been used in many learning areas across various disciplines. Brookhart & Chen (2014) pointed out a number of advantages of rubric marking such as ensuring reliability between raters by reducing subjectivity, bridging summative and formative assessment forms, and offering criterion-level diagnostic information.

Various rating tendencies such as the halo effect, rater leniency, restriction of range and central tendency have been considered affecting rubric-based

assessment (Myford & Wolfe, 2003). It was proved by Humphry & Heldsinger (2014) that grid-like, common structure of rubrics causes local independence violations across ratings as each criterion possesses common-numbered categories, which weakens validity by restricting construct-relevant variation in scores. If a rubric was not employed effectively by raters as raters are not well trained, their assessments would be biased. This is about high-stakes, large-scale standardised writing assessments since it takes time and resources to train assessors successfully (Rezaei & Lovorn, 2010).

On the other hand, development of scales and descriptors for each level has a crucial part for the assessment validity, as McNamara (1996) stated. Scales shape the notion of what abilities or skills are intended to be measured by the test. That's why, while determining a scoring system, one of the first decisions to be made is what type of scale should be used (Weigle, 2002). Three rating scale types are presented below:

Primary Trait Scoring. This scoring type came out in an attempt to guide a large-scale testing program in the US by Lloyd-Jones (1977). Understanding how well learners are able to write in a narrow discourse such as explanation or persuasion is the idea behind this type. The scale is determined to be corresponding with the assignments. For each writing task, a scoring rubric involves: (1) an explanation about why it was scored; (2) the writing task; (3) statement of primary rhetorical trait; (4) statement for the relationship between the primary trait and task (5) at each level, sample scripts; (6) a rating scale articulating performance levels; (7) a hypothesis on the desired task performance (Lloyd-Jones, 1977). The rubric is quite detailed in terms of learners' different perspectives and approches about the task, which explaines why it is labor- and time-intensive, according to Lloyd-Jones (1977).

Holistic Scoring. Holistic scoring, assigning just one score to a task grounded on the overall impression of it, is used by many assessment programs. Each task is read and judged quickly against a scoring rubric outlining the scoring criteria. In the rubric, at each level there are a set of benchmarks or anchor scripts exemplifying the criteria, and while scoring tasks, raters keep to that rubric (Weigle, 2002).

Holistic scoring has become popular and been used in assessment over the past 25 years. Raters do not need to read tasks several times and focus on different writing aspects in each time. It is fast, so raters read a task once and assign a score immediately (Weigle, 2002). As one of the adherents, White (1984) specified a few advantages of this type of scoring. He emphasized that holistic scoring has students concentrate on the strong sides of their writing, not its deficiencies. Therefore, it brings a reward system for the writers doing well. As students focus on definite aspects of writing and what is most necessary in the context, it informs them about those aspects in a prolific manner. White (1984) also stated that it is more valid than analytic one since readers' most personal and authentic reaction to the tasks is reflected in holistic scoring.

As for the disadvantages of holistic scoring, the biggest one is that assigning a single score does not give much diagnostic information about students' writing abilities on different aspects of writing such as organization, control of syntax, vocabulary usage, and so on. As different aspects improve at different rates, this case is especially difficult for second language learners. Another drawback is that it is not easy to interpret holistic scoring because the same criteria are not necessarily used by all raters to get to the same scores. To illustrate, a rater can give 4 to a task on a holistic scale because of its development, organization, content features whereas another one gives 4 to the same task because of its control of vocabulary and grammar features (Weigle, 2002).

Analytic Scoring. In analytic scoring, tasks are not given a single score. On the contrary, they are rated on criteria or several features of writing such as cohesion, grammar, mechanics, content, register, organization or vocabulary, thus presenting detailed information about students' performance in various aspects of writing, which is the reason why it is preferred by many writing specialists (Weigle, 2002).

Compared to holistic scoring, analytic one has a few more other advantages, as well. First, in rater training, analytic scoring is more advantageous since inexperienced raters have the opportunity to apply and comprehend the criteria easily in separate scales (Weir, 1990). Second, for second language learners having an uneven profile across various features of writing, analytic

scoring seems more appropriate. Finally, reliability increases when a discrete-point test is added additional items, so analytic scoring is more reliable than holistic one as each task is given multiple scores (Hamp-Lyons, 1991).

As for disadvantages of analytic scoring, the biggest one is that it is time-consuming as the raters are supposed to make more than one decision for each task. An extra problem with it is; provided that scores on different scales are intermingled in order to come up with a composite score, the information gained about students' scores by the analytic scale can be lost. It can also be possible that if scores are intermingled into a single score, experienced raters in analytic scoring can rate more holistically (Weigle, 2002).

As emphasized above, large-scale assessment programs have become dominant across the world. To provide the accountability they require, standardised writing assessments have become an important asset. More research studies are needed to resolve the problems and disadvantages of standardised writing assessment and scale types.

Standardisation Studies on Writing Assessment. While assessing writing, the reliability of the scores should be ensured. Studies before the 1990s were basically there to find measures to achieve this. Those measures included improvement in rating scales' design and standardisation studies such as moderation sessions and marker training. In training sessions, markers are shown how to use the scale by checking benchmark samples and this leads to the standardisation of scoring writing (Hamp-Lyons, 1991).

To ensure reliability, the following measures are suggested by Hoghes (2003):

- Create an appropriate scale
- Calibrate the scale
- Train markers
- Follow scoring procedures

Weir (2005), on the other hand, emphasizes the importance of rating scale, markers, rating procedures, grading and awarding. Standardisation, marker training, moderation, statistical analysis of the scores and rating conditions are the variables of the rating procedure. These together increase the validity rate, as well.

To have standardisation in the scoring of writing, various systems have been acknowledged as successful such as explicit scoring rubrics (Linn & Gronlund, 2000), a fairly long scoring scale (Coffman, 1971), moderation or cross-checking of marking (Gronlund & Linn, 1990), expert markers and augmentation of holistic grades (Penny, Johnson & Gordon, 2000) and systematic scoring processes (McMillan, 2001).

One of the most important attempts on getting a standardized assessment on writing was by the New Standards Project in New Zealand in the early 1990s. Each state had the chance to use their own scoring rubric. In the project, a pilot study was carried out with the usage of a seven-point scale. 114 teachers scored three writing tasks by using the rubric in question. The exact agreement percentages were between 40%-49% with the adjacent scoring percentages between 86%-88%. Yet, consistency coefficient average was only .54 across those writing tasks (Resnick & De Stefano, 1993).

A related work in New Zealand was reported by Gearhart, Herman, Novak, Wolf, & Abedi (1994). The scoring of writing by three raters was compared. Six-point analytic scoring rubrics were made use of, and the results for the exact agreement were 39%-46% and 28%-37%, the results for the adjacent agreement were 95%-97% and 92%-94%. As for the consistency correlations, they ranged from .48 to .68.

In Rochester, New York, an eight-point rubric was used by 20 trained raters and classroom teachers to assess the writing portfolios of K-2 students. It was found out that exact agreement consensus was obtained between the raters, which is 63%-73% and the consistency coefficients were between .68 and .73 (Supovitz, MacGowan & Slattery, 1997).

In another similar study in Pittsburgh, Pennsylvania by LeMahieu, Gitomer, & Eresh (1995), a district-wide portfolio assessment process with a six-point rubric used by 25 trained raters was carried out with 12. grade. In this study, consensus rates were achieved by having adjacent agreement of 87%-98% with consistency coefficients between .75 and .87.

In a comparative study carried out in Louisiana in a pool of trained and untrained raters using a four-point rubric with six categories, measurement coefficients were between .61 and .66 (Stuhlmann, Daniel, Dellinger, Denny & Powers, 1999).

In United Kingdom, Green and his-coworkers (2003) reviewed the consistency of scoring between the years of 1995 and 2002 at Key Stage level since high-stakes writing assessment was comprehensively used. Children's writing samples from the same school were scored by both their classroom teacher and central authority. Consistency correlations in 1995 were .59 and .69 while in 2002 they were .57 and .71.

Obviously, it is a hard job to have exact egreement even we use short scoring rubrics, and so the consensus values become more robust. In the studies reviewed above, consensus estimates are between 40% and 60%, adjacent agreement 80%-100%, consistency coefficients between .70 and .80 in the standardized assessments in writing (Glasswell, Brown & Harland, 2004).

Another important component that should be mentioned here is the training of markers before they start scoring writing tasks of students. With this thought in mind, a half-day training program was provided by the University of Auckland for seventeen experienced teachers none of whom had had any experience in large-scale assessment procedures. To ensure the quality, the teachers checked a number of scripts and the cross-checking of scoring was done by expert markers. The training included a 1.5-hour of grammar instruction lecture and an overview of the rubrics. Nearly 15-20 minutes were spared to clarify the rubrics to be used and the tasks before each and every writing task was started (Cooper, 1984).

As for marking, sample scripts were provided earlier to be discussed, which also served as a standardization study. When all the tasks were completed, following training on rubrics took almost an hour. In sum, training time for this scoring study was 4 hours (Cooper, 1984). 7.2 scripts per hour were averaged by the markers. Cross-checking was carried out by an expert marker to assess consensus between markers' scorings. After the cross-checking, all markers were provided with feedback on their marking. After the study, grammar and language

resources such as punctuation, complex sentences and sentence structure were ascertained to be the areas that the teachers needed extra instruction in. (Alderson, Clapham & Wall, 1995). Consensus coefficient was .75, consistency coefficient was .75 and lastly, measurement coefficient was .77 in average (Glasswell, Brown & Harland, 2004).

The CEFR A2 Level for Writing Skill. The CEFR Levels provide a basic 'global' representation by which non-specialist users can communicate to the system, and curriculum planners as well as teachers have orientation points (Alderson, 2007)h. Within this context, it is stated in the global scale that an A2 level language learner:

> "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); communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters; describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need" (Council of Europe, 2001, p. 24).

In order to have learners and other users oriented, a grid that indicates language use categories at each level is presented as an overview. This self-assessment orientation device assists language learners to have their own language skills profile. It has come to be the major source in planning, not only for writing production but also for different skills. It has become easy for teachers, learners and course designers to determine course objectives, organize content, specify what students are supposed to do after completing a course or a study program. As for A2 level, it goes as follows in the grid.

> "I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions." (Council of Europe, 2001, p. 26).

In the following parts, checklists, written text types and the features of A2 level are explained in detail to shed light on writing skill in the CEFR at A2 Level.

The CEFR Checklists. The CEFR describes in an extensive way what learners should learn to do so as to use a language to communicate and what kind

of skills and knowledge they should develop in order to act effectively (Council of Europe, 2001). As it is implied in these words, the description's major orientation is behavioural: all the activities learners deal with and the tasks they carry out when they speak, listen, write and read in a foreign language. This behavioural orientation can be the most crucial innovation of the CEFR. Similarly, "can do" descriptors can be made use of to describe teaching and learning curriculum and plan, by which the CEFR has curriculum, pedagogy and assessment be in a closer relation with each other than the traditional case (Little, 2009, pp. 1-2).

The checklists of 'I can' descriptors in the language biography are designed in accordance with the proficiency levels as well as communicative activities specified in the CEFR. Learning goals and self-assessment are identified by the checklists. (Council of Europe, 2018). They not only assist learners and teachers to embrace a task-based orientation to their learning and teaching process, and they also help learners and teachers plan, monitor and assess their own learning (Little, 2006, p. 184). To have such checklists is also a key enabling portfolio-based assessment, and language learning gains a new dimension. They make it easier to link instruction to assessment, to assist learners to have learner autonomy and to take responsibility for learning, to promote reflection and to have learners take risks (Ekbatani, 2000, pp. 6-7).

The purpose of checklists is to develop performance-related scales that describe what learners can literally do in the target language. They are user-orientated and help non-specialists have a better communication concerning the testing process and the interpretation of the test results. They provide;

a) a functional tool for the ones involved in language teaching and testing. They help define what stage the language users are at and what they can do.
b) support for developing and preparing diagnostic test tasks and teaching materials.
c) a means of activity-based linguistic examination that can be used while recruiting or training people in terms of language teaching.
d) a useful tool to compare different languages in terms of their course and material objectives (Council of Europe, 2001).

To have a better understanding on checklists -especially the ones on written products, the tables below by Lenz and Schneider (2004) are of great help:

Table 4

A2 Level Self-Descriptors for Overall Written Production

OVERALL WRITTEN PRODUCTION
Can write a series of simple phrases and sentences linked with simple connectors like „and", „but" and "because". [CEFR-2001]
(Taken from "A bank of descriptors for self-assessment in E.L.Ps 2004", p. 16)

Table 5

A2 Level Self-Descriptors for Creative Writing

CREATIVE WRITING
Can write about everyday aspects of his/her environment, e.g. people, places, a job or study experience in linked sentences. [CEFR-2001]
Can write very short, basic descriptions of events, past activities and personal experiences. [CEFR 2001]
Can write a series of simple phrases and sentences about their family, living conditions, educational background, present or most recent job. [CEFR 2001]
Can write short, simple imaginary biographies and simple poems about people. [CEFR 2001]
(Taken from "A bank of descriptors for self-assessment in E.L.Ps 2004", p. 19)

Table 6

A2 Level Self-Descriptors for Overall Written Interaction

OVERALL WRITTEN INTERACTION
Can write short, simple formulaic notes relating to matters in areas of immediate need. [CERF 2001]
(Taken from "A bank of descriptors for self-assessment in E.L.Ps 2004", p. 78)

Table 7

A2 Level Self-Descriptors for Correspondence

CORRESPONDENCE
Can write very simple personal letters expressing thanks and apology. [CERF-2001]

(Taken from "A bank of descriptors for self-assessment in E.L.Ps 2004", p. 81)

Table 8

A2 Level Self-Descriptors for Notes, Messages and Forms

NOTES, MESSAGES & FORMS
Can take a short, simple message provided he/she can ask for repetition and reformulation. [CEFR 2001]
Can write short, simple notes and messages relating to matters in areas of immediate need. [CEFR 2001]

(Taken from "A bank of descriptors for self-assessment in E.L.Ps 2004", p. 83)

Table 9

A2 Level Self-Descriptors for Orthographic Control

ORTHOGRAPHIC CONTROL
Can copy short sentences on everyday subjects – e.g. directions how to get somewhere. [CEFR 2001]
Can write with reasonable phonetic accuracy (but not necessarily fully standard spelling) short words that are in his/her oral vocabulary. [CEFR 2001]

(Adapted from "A bank of descriptors for self-assessment in E.L.Ps 2004", p. 96)

A2 Level (Waystage). Waystage is specified as an important direction in terms of the planning phase of learning activities since it serves as an early language learning objective. Rather than demotivating the learners by requiring them to do something higher than their potential, a less demanding and challenging learning load was seen appropriate (Bariskan, 2006, p. 49). The learning load that the learner should undergo in Waystage level is just the half of Threshold Level, which shows that a learner can reach Threshold Level by

studying on the language in around 200 hours, including proper guidance and individual work, of course (Van Ek, & Trim, 1998, p. 4).

To put it another way, Waystage is a service stage that leads all the way to Threshold because the basic components of this higher level are the premises of Waystage. Considering this, it can be said that those who are keen on having basic English acquisition are the ones Waystage was designed for. In other words, they are defined as the general beginners; they ask and answer questions, state an opinion about present, past and future events, give reasons for a situation and understand what the situation is about. In this regard, Waystage is the most suitable objective for the ones who desire to learn English for communication purposes (Bariskan, 2006, p. 50).

At Waystage; the learners, as speakers, are expected to make themselves understood not just by native or near-native listeners, but also by other non-natives that have reached Waystage. As for the learners as listeners, they are expected to understand what they are told, not just by native or near-native speakers, but also by other non-natives that have reached Waystage. What has been stated about the learners as speakers is the same for the learners as writers. Yet still, they probably aspire to satisfy the criteria of formal correctness, specifically in letters that they write to strangers. Accessible appropriate reference works give them the chance to satisfy such criteria easily when compared to the spoken language production. All these suggest that co-operation and tolerance are what the learners' communication partners are expected of (Van Ek & Trim, 1998, p. 67).

As stated by Van Ek (1991), the learner at Waystage, as a reader, can:

- figure out complex words and their meanings which the learner is familiar with and which are combined by taking word formation rules into consideration.
- figure out unfamiliar words and phrases and their meanings from a context featuring familiar elements to make the meanings in question identified.
- figure out so-called 'international words' that the learner is familiar with from his mother tongue; the learner is aware of the meaning differences in 'international words' and connects words of two different languages.

- use a billingual dictionary or a word list that is arranged alphabetically to find unknown words' meanings (Van Ek & Trim, 1991, pp. 56-59).

The learner at Waystage, as a listener, can:

- perform the operation explained in the first clause above if this process does not require any further phonological changes other than linking and vowel/consonant changes in the constituent elements.
- perform the operation explained in the second clause above if contextual clues are provided in a way that makes them interpretable without entailing context backtracking and reconsideration.
- perform the operation explained in the third clause above if the phonological differences that the foreign and the native language forms have are limited to the standard correspondences these two languages have in common.
- obtain information on a specific topic from a context that involves unknown elements (Van Ek & Trim, 1991, pp. 56-59).

The learner at Waystage, as a speaker, can:

- introduce a rephrasing.
- make use of paraphrase, especially general words or a superordinate to describe while indicating colours, sizes and shapes of general physical properties.
- make use of qualities and properties to describe.
- identify by indicating.
- appeal for assistance.

The learner at Waystage, as a writer, can:

- express ignorance.
- make use of the devices mentioned above; paraphrase, especially general words or a superordinate to describe while indicating colours, sizes and shapes of general physical properties as well as qualities.
- use both bilingual and monolingual dictionaries of an appropriate kind (Van Ek & Trim, 1991, pp. 56-59).

The learner at Waystage, as a social agent, can:

- apologise for his/her behaviour.

- talk about customary things in his/her country.

- ask for guidance.

The learners at Waystage are supposed to use the techniques and strategies above as well as other privileged devices they wish to use. Techniques such as finding specific information in reference works or grammatical surveys, and strategies such as using synonyms for unknown words, experimenting with word formation, using grammatically imperfect forms can be, not necessarily, included by the learners. Those who provide learning facilities are to decide on which of these devices can be adopted by the learners (Van Ek & Trim, 1991, pp. 56-59).

As for the objectives, according to Van Ek and Trim (1998), the following general purposes are expected to be used by the learners who complete A2 Level-Waystage:

1. performing certain transactions

    a. being able to make travel and accommodation arrangements and appointments, etc.

    b. making purchases

    c. ordering food/drink

2. giving and getting factual information

    a. personal information (name, nationality, address,etc.)

    b. non-personal information (about facilities, places, rules, services, how to get there and where to eat, etc.)

3. establishing social or professional contacts

    a. meeting people

    b. making and responding to invitations

    c. arranging a course of action

    d. exchanging information, feelings related to personal life, environment, interests, leisure activities, etc (Van Ek & Trim, 1998, p. 10).

Other than the purposes, the learners at Waystage depend largely on their ability to cope with the particular themes or topics that they would be likely to

handle. There are many different ways to classify thematic categories. According to Van Ek and Trim (1998), these themes are:

a) Personal Identification

The learners at Waystage can talk about themselves including their age, gender, job, nationality, family, address, telephone number, marital status, where and when they were born as well as their religion, likes and dislikes. They can also get information from others on similar topics.

b) House and Environment

The learners can describe and talk about a house with the rooms in it including furniture and services as well. They can also obtain similar references and descriptions from others.

c) Daily Life

The learners can describe and talk about their daily life at home or work; get information from others and exchange views on the same issues.

d) Free Time / Entertainment

The learners can describe what they do in their free time and state their hobbies and interests, sports, public entertainment and reading. They can also use entertainment facilities, and get information from others on the same issues.

e) Travel

The learners can book tickets, use means of public transport and the road traffic system, arrange holiday trips and accommodation for travellers, deal with travel forms and documents and exchange information with others on these issues.

f) Relations with Other People

The learners can take part in social life, refer to personal relations and handle correspondance matters.

g) Health and Care

The learners can state if they feel sick or well; refer to matters of health, comfort, hygiene and illness; state what is wrong with them to a doctor and report accidents and exchange information with others on these issues.

h) Shopping

The learners can make use of shopping facilities, especially getting food, clothes and household items; pay for their purchases and exchange information with others on these issues.

i) Food and Drink

The learners can order food and drinks in restaurants and cafes. They can also exchange information with others on these issues.

j) Services

The learners can make use of telephone, bank, postal and medical services, police, petrol stations and car maintenance services.

k) Places

The learners can ask for directions and give directions to the strangers.

l) Language

The learners can handle the problems on understanding and expression as well as refer to language ability.

m) Weather

The learners can get information from weather forecast programmes and exchange information with others on these issues (Van Ek & Trim, 1998, pp. 30-42).

Written Text Types for A2 Level. According to Van Ek (1998), learners are supposed to produce the text types below at Waystage (Van Ek & Trim, 1998, p. 45):

a) required data in forms:
  - hotel registration forms (check in, check out, visa application, bank account and job application, etc.)
  - forms required when entering or leaving a country
b) standard letters:
  - booking accommodation
  - simple letters and postcards
c) personal correspondence:
  - simple messages such as greetings and congratulations
  - simple private letters concerning matters of common interest to themselves and friends or acquaintances.

d) short messages and notes

e) greetings and congratulations

f) written directions

g) written instructions

h) descriptive paragraphs

i) written dialogues

Assessing Written Products. While assessing written products of the learners, the first step is to analyze the effects of tasks, raters, rating criteria and learners on the learners' written responses in terms of the variability in the ratings. The second step is to investigate if the analyses in question provide empirically grounded scores that are in accordance with the CEFR with regard to the proficiency levels of the tasks ranging from A1 and A2 Levels as the basic user stage, B1 and B2 Levels as the independent user stage to C1 and C2 Levels as the proficiency user stage (Harsch & Rupp, 2011).

On the other hand, regarding the use of the CEFR to assess writing, Harsch (2007) stated that the scales of the CEFR are too indefinite, and sometimes unclear to be used for writing task development or rating scales. To get rid of this drawback of the CEFR, a grid for classifying writing tasks was developed by ALTE (the Association of Language Testers in Europe) for the CoE (2008). The purpose is to analyse test task content and other attributes, facilitating comparison and review to promote the specification while aligning language tests to the CEFR. After that, the CoE (2009) published "the Manual for Relating Language Examinations to the CEFR"

Until recently, empirical studies on the relation between language acquisition and language testing perspectives have not been enough, but with the appearance of the CEFR, there has been an interest in carrying out studies putting language testing perspectives and language acquisition together across Europe (Alanen et al. 2010, pp. 22). Language testing requires correct and reliable measures in order to assess language proficiency or communicative language ability. It is basically concerned about the success of the items which are made use of in language testing as well as in tasks' communicative adequacy, and few studies exist making use of qualitative ratings to assess adequacy. The CEFR

scales, as a way of qualitative rating, are appropriate to be used to assess adequacy in open tasks (Pallotti, 2009). Since it is essential to take advantage of proficiency scales to assess communicative adequacy, the CEFR scales are of great importance to provide descriptors in order to assess language performance features relevant for a specific proficiency level (Alanen et al. 2010, p. 83).

The term "language proficiency" in the CEFR is described in terms of function ('can-do statements') which determines domains and functions that learners cope with in the language, and with regard to quality which describes how well language use is (Hulstijn, 2007). As for foreign language writing proficiency, it has been a term that went through several measures to engage aspects of foreign language in writing, which are basically accuracy, fluency and complexity. Several studies have carried out on writing proficiency by considering data from language tests like IELTS and ESOL examinations by Cambridge (Banerjee et al. 2004), and some of them have a direct link to language testing (Alanen et al. 2010, p. 28).

Designing and Selecting Communicative Writing Tasks. The writing ability is generally assessed by open tasks that a set of written responses can be obtained from. And these are graded by trained raters with the help of a rating scale covering proficiency levels. This approach is called a 'multilevel approach'. Yet, if wanted to know one specific level that learners have reached, an approach in which tasks targeted at a particular level are used; a fail/pass rating instrument which is used to assess learners' written responses should be made use of. This approach is called a 'level-specific approach' (Harsch & Rupp, 2011, p. 3).

As a key point connecting research and language testing, task plays a great role in both measurement and data elicitation. It is also crucial in performance-based assessment of language proficiency. As Brindley (2009) puts it, task-based language assessment is;

> "…the process of evaluating, in relation to a set of explicitly stated criteria, the quality of the communicative performances elicited from learners as part of goal-directed, meaning-focused language use requiring the integration of skills and knowledge" (Brindley, 2009, p. 437).

With regard to writing tasks' alignment, the Manual proposes to identify the tasks by using formal standard-setting methods along with the grid. In the area of

writing tasks, 'examinee-centered standard-setting methods' are recommended, using examinees' responses to connect the CEFR levels to the writing test (Council of Europe, 2009) . Yet, should one want to connect the writing tasks to the CEFR levels themselves with 'test-centered standard-setting methods', the Manual does not offer one for writing tests. (Harsch & Rupp, 2011). Even though this thesis does not handle the formal standard setting procedure subject, its purpose is to investigate to what extent the analyses carried out in it can contribute to underpin the alignment of the CEFR with writing tasks with empirical scores.

It is for sure a demanding job to design and select communicative writing tasks appropriate for both language acquisition and language testing. They should provide data needed to see the differences in linguistic features, and enable to assess task performances. In language testing, various tasks or tests for beginner, intermediate or advanced level learners are used for this purpose (Alanen et al. 2010, p. 30).

Alanen and her colleagues (2010) proposed a number of solutions to relate language acquisition to language learning perpectives while designing and selecting writing tasks:

- asking learners to carry out all task types irrespective of their age, level or trying to link tasks with their ability,
- considering the scales like the CEFR to be used for writing assessment,
- considering the proficiency level that tasks are aimed at, tasks' topics and domains and the language functions expected to be used in the tasks,
- considering specific linguistic structures that tasks elicit (e.g. verb forms, questions, negation, locative expressions, etc.) in order to ensure task adequacy and communicative authenticity,
- paying attention to task types and processes to be cummunicative and to have authenticity,
- piloting tasks to yield information (Alanen et al. 2010, pp. 31-32).

Written Products and Their Features. Writing is a valid way of language expression and it offers additional activities that learners make the most of beyond the classroom. It helps learners improve their other skills, memorise language and clarify meaning by clarifying the spoken language. By means of writing, learners focus on accuracy and target language aspects that are not obvious in the spoken language all the time (Pachler & Redondo, 2006). Herein, it is fair to say that writing is a process where not just the 'finished product' is focused on, but the writer as well (Macaro, 2003). Besides, the language used in the written products tend to be standardised as it is less flexible than other skills (Broughton Et al. 2003: p.116).

In the following part, written products and their features are explained to have a better understanding on the impact of writing in language learning. Yet, it will be necessary to look through the process of writing first.

The Process of Writing. Good writing means more than sitting down and writing about something on a piece of paper. It includes thinking, planning, writing, revising (Houge, 2007, p. 28). Therefore, writing process involves several steps:

a) *Pre-writing:* This is the part that the learner thinks, talks and organizes his/her ideas before writing the first draft. After that, he/she brainstorms ideas before choosing which idea to write about first, next and last.

b) *Drafting:* In this part, the learner creates an outline, clusters the ideas he/she got in pre-writing stage into logical parts and writes his/her ideas into full sentences.

c) *Editing:* The learner checks what he/she has written, adds more information if necessary, corrects any mistakes and revises the text in terms of content and structure.

d) *Writing a final draft:* The learner rewrites the draft in order to make it look neat, proofreads it to check spelling and grammar and, makes final corrections (Zemach & Rumisek, 2005, p. 3).

In writing process, it is important to take the written products and their features into consideration as follows. These types of written products are of a great place in the study as they are the ones the students in the focus university are requested to write for this study.

Paragraph Writing. A paragraph is a set of related sentences on a single topic which is just one idea that is dealt with in the paragraph. And a paragraph has three parts (Houge, 2007, p. 4):

1. **Topic sentence**

    The first sentence in the paragraph names the topic and states what the paragraph will be about. We call this sentence the topic sentence. It has two parts: a **topic** and a **controlling idea.** The topic part is the part naming the topic. The controlling idea, on the other hand, states what the paragraph says about the topic. Generally, the topic comes first, and the controlling idea second in the topic sentence (Houge, 2007, p. 48).

    For example;

    English is useful to find a good job.

    Topic   controlling idea

2. **Supporting sentence**

    Supporting sentences, the middle sentences, give details and examples on the topic. They are the biggest part in a paragraph.

    While writing supporting sentences, the rule of **unity** must be followed; all sentences in a paragraph should be about just one main idea. One way of this rule is **relevance**. All sentences should be relevant.

3. **Concluding sentence**

    Concluding sentence is the last sentence in a paragraph. It closes the paragraph in order not to leave the reader to expect more.

    - Sometimes it reminds the main points to the reader by using different words to state the topic sentence again.
    - Sometimes it summarizes the main points (Houge, 2007, p. 48).

Considering paragraph writing, one of the most important point is **"outline".** It helps organize ideas before beginning to write. It is like an architect's plan designing a house. Besides, you make sure that nothing important is left out. A detailed paragraph outine is as follows:

Table 10

Paragraph Outline

Topic Sentence

    A. Main Point (reason, benefit, and so on)
        1. Supporting Sentence (example, fact, description, and so on)
        2. Supporting Sentence
        3. Supporting Sentence
    B. Main Point
        1. Supporting Sentence
        2. Supporting Sentence
        3. Supporting Sentence
    C. Main Point
        1. Supporting Sentence
        2. Supporting Sentence
        3. Supporting Sentence

Concluding Sentence

(Adapted from Oshima & Hogue,2007)

Now, it is necessary to look at the paragraph types and their features.

    Narrative Paragraphs. A narrative paragraph is the one telling a story. It has a beginning with a development and an end part. The writer is free to write about a journey, a holiday or an important event in his/her life. The story does not necessarily have to be personal; it can be about the stories of other people (Koc et al. 2008, p. 78).

    The points to be considered while writing a narrative paragraph are as follows:

- ***Coherence and Cohesion:*** Coherence means putting all sentences into a correct order. Chronological ordering is a must in a narrative paragraph. What this means is that the writer should write the events according to the time starting from the beginning leading to the end. Cohesion, on the other hand, means having well-connected sentences in the paragraph. In order to achieve this, cohesive devices should be used (e.g. linking words, the definite article, synonyms, etc.)

- ***Unity and Completeness:*** In a narrative paragraph, having all the supporting sentences related to the topic sentence brings unity, which makes it easy for the readers to understand. Completeness, on the other hand, means explaining all the supporting ideas fully. There should not be any missing part (Koc et al., 2008, p. 78) (see Appendix A).

Descriptive Paragraphs. As the name suggests, descriptive paragraphs are the ones describing something. This description can be about a person, an animal, a place or an object. The important thing about writing a good descriptive paragraph is that the writer should be able to describe each and every detail in order to create a mental picture of it for the reader. Many adjectives, prepositions should be included in the paragraph to give information on the location, size, color, appearance and the spatial ordering. Five questions relating to the five senses should be asked by the writer before starting to write: *What does it look / sound / smell / taste / feel like?* (Koc et al., 2008, p. 90)

In the organization part of descriptive paragraphs, there seem to be two key points that should be taken into consideration:

1. **Using space order in order to organize the description.**

   To get a better understanding on this, imagine your standing in your classroom's doorway. How would you describe this place to someone not being there? Two possibilities are in question: you would either start describing the left side of the doorway and continue in a clockwise direction, then end at the doorway again, or start describing the front of the room and move from front to back. This is called **space order.** And other kinds of it are (Houge, 2007, pp. 98-101):

   - top to bottom   - bottom to top   -right to left   -left to write

   - far to near   -near to far   -outside to inside -inside to outside

2. **Using a lot of descriptive / specific details.**

   Writing a descriptive paragraph is like painting a picture with words. The reader should see what you have tried to describe. Using specific details helps you in this point. The more specific you are, the better readers see it (see Appendix B). Some examples are as follows:

Vague: a lot of money,

a big house,

a nice car.

Specific: £ 430,000

a four-bedroom house

a Porsche (Houge, 2007, p. 101)

Types of Assessment

Anyone working in language teaching can make use of the CEFR as a consciousness-raising instrument for assessing language ability with the help of calibrated scales. To put it differently, the CEFR can be used as a solid basis for language test development, but it should not be perceived as a how-to guide. Therefore, most test developers struggle to design tests in accordance with the CEFR, not just in philosophy, but also in practice (Harsch & Rupp, 2011, p. 2) . With the CEFR's final version published in 2001 in the European Year of Languages, (Scharer, 2000) it has had a huge effect on educational systems, and had language testing agencies associate their tests and assessment types with the CEFR levels (Little, 2009). In the list below, we have assessment types that can be related to the CEFR, and there is no significant distinction between the assessment types in terms of their placement on the left or right (Council of Europe, 2001, p. 183).

Table 11
Types of Assessment

1	Achievement assessment	Proficiency assessment
2	Norm-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 judgement
11	Holistic assessment	Analytic assessment
12	Series assessment	Category assessment
13	Assessment by others	Self-assessment

Achievement Assessment / Proficiency Assessment. Achievement assessment means assessing what has been taught – assessing the attainment of some specific objectives. That is why, it is a part of the syllabus, the course book and the week's/term's work. As teachers want to get feedback for their teaching, they have an intention to be more into achievement assessment. Needless to say, it is more about learners' experiences. It also has a proficiency angle as it assesses language use in relatable situations and the purpose of giving a picture of competence that is emerging (Council of Europe, 2001, p. 183).

As for proficiency assessment, it represents what learners know and can do in terms of applying the subject to the real world. While achievement assessment embodies an internal perspective, proficiency assessment embodies an external one. That is why, adults and employers prefer proficiency assessment as they want to see the result of outcomes, and what they can do. That is one advantageous side of proficiency assessment: it is there to assist people to see where they are, and results are totally transparent. The communicative tasks in this type of assessment help learners to see what they have accomplished, which has an achievement element (Council of Europe, 2001, pp. 183-184).

Norm-Referencing (NR) / Criterion-Referencing (CR). While placing learners in a rank order with regard to their peers, norm-referencing is the topical issue. The class or demographic crowd taking a test can be effective in norm-referencing. It is normally used to form classes by placement tests. While designing a norm-referenced test, experts compare raw scores they get from the people who have taken the new test with the ones they get from people that have taken previous tests. This process is preferred so as to create "expected scoring patterns for the test" (Wright, 2007, p. 14). Standardisation can be carried out where data are collected from some specific population types that have attained learning outcomes in question. By standardisation, it means the test is uniform, all people taking the test cope with the same tasks. Administration and scoring of them are done in the same way (Koreltz, 2008, p. 23). This procedure is performed by assessment designers who also determine the scores for average, below and above for the specific objective being tested (Isaacs et al. 2013 pp.97).

In Criterion-referencing, on the other hand, learners are assessed with regard to their ability on a specific subject – regardless of their peers, which makes it against norm-referencing. The purpose of this assessment type is to have the results of the learners on a specific test placed with regard to the total criterion space. This includes: "(a) the definition of the relevant domain(s) covered by the particular test/module, and (b) the identification of 'cut-off points': the score(s) on the test deemed necessary to meet the proficiency standard set" (Council of Europe, 2001). Criterion-referenced scores involve their potential to draw a distinction between the learners who have or have not improved their skills or abilities that are essential for a particular activity or field (Coaley, 2010). They are carried out in educational institutions where students' success at a specific level is specified by a framework of predetermined standards (Isaacs, T., et al. 2013 pp. 41).

Mastery CR / Continuum CR. The mastery criterion-referencing is a kind of assessment approach in which degrees of quality in the attainment of the objectives are not given importance, but learners are divided into categories as 'masters' and 'non-masters' with a single competence standard. The mastery approach is actually an achievement approach that is about the course or

module's content, and placing achievement on the continuum of proficiency is not of much importance (Council of Europe, 2001, p. 184).

In the *continuum criterion-referencing approach,* a learner's ability is assessed with a pre-determined continuum of all related degrees that are in question. It is a kind of alternative to the mastery approach. The criterion is that continuum, which makes it certain that the results of the tests are important to be taken into consideration. A scalar analysis like Rasch model can be made use of for referencing to this criterion with the aim of making connection between the results from all tests (Council of Europe, 2001, p. 185).

The Framework can be put to use with mastery or continuum approach. The Common Reference Levels can be matched to the continuum approach and the scale of levels that are used in it. The Framework can assist the mastery approach by offering levels and grid of categories to map the objectives considered to be mastered (Council of Europe, 2001, p. 185).

Continuous Assessment / Fixed Point Assessment. Continuous assessment is done by the teacher based on the class performances, works and projects done by learners throughout the course time. Accordingly, the final grade is determined based on the whole course time. One of the assessment types of the writing tasks worked on in this study is continuous assessment, as well. Assessment is merged with the course and assists the end-of-course assessment as it did with the tasks carried out in the study. Continuous assessment can be in checklists, grids and forms that are filled in by both teachers and learners as well as homework and regular achievement tests. It can also lead to creativity and various strengths, but its being teacher-dependent becomes a burden for teachers as it requires bureaucratic procedures (Council of Europe, 2001, p. 185).

Teachers, trainers and lecturers all assess the on-going curriculum process in the classroom as well as the learning environment. Professionals also get the opportunity to evaluate the teaching strategies that they implement in the curriculum, and make changes in those strategies by taking learners' responses into account. Besides that, continuous assessment gives information about the attainment of some specific skill levels, not marks or scores. Learners monitor their

own improvement, and peer assessment is easily incorporated (Isaacs et al. 2013, p. 34).

Fixed point assessment involves the time when an examination, besides other assessment, is completed and the grades on them are awarded at the end or before the beginning of the course. In this type, what is important is what the individual can do now, not what he/she has done earlier. Assessment is considered to take place at fixed points so as to reach a decision. This type of assessment ensures that learners can still achieve things that were on the syllabus long time ago, but at the same time it may cause examination traumas for some learners (Council of Europe, 2001, p. 185).

Formative Assessment / Summative Assessment. Formative assessment is a never-ending assessment process to gather information on learners' strengths and weaknesses in order to use the information to plan their course, which is also preferred at the SFL concerning the tasks performed by the students. Questionnaires and consultations that provide non-quantifiable information are also involved in formative assessment. The purpose is to improve learning, but its weakness is in the feedback part. Feedback functions "if the recipient is in a position (a) *to notice*, i.e. is attentive, motivated and familiar with the form in which the information is coming, (b) *to receive*, i.e. is not swamped with information, has a way of recording, organising and personalising it; (c) *to interpret*, i.e. has sufficient pre-knowledge and awareness to understand the point at issue, and not to take counterproductive action and (d) *to integrate* the information, i.e. has the time, orientation and relevant resources to reflect on, integrate and so remember the new information" (Council of Europe, 2001, p. 186). This is basically about self-direction which suggests the individual monitoring of their own learning as well as finding out ways of working on feedback (Council of Europe, 2001).

An assessment activity can be seen as a tool helping learning on condition that it supports teachers and students by providing information as feedback in assessing each other or themselves. Such assessment takes formative assessment style when the evidence is put to use in an effort to link the teaching work to fulfil the learning needs (Black et al. 2003, p. 2).

According to William's chart for the evolution of formative assessment, it gives information on;

- the learning process
- the learning process which teachers make use of for instructional decisions
- the learning process which teachers make use of for instructional decisions and also learners make use of it in improving their performance
- the learning process which teachers make use of for instructional decisions and also learners make use of it in improving their performance that motivates learners in return (Wiliam, 2011, p. 8).

Summative assessment, as the name suggests, summarizes the attainment of objectives at the end of the term/course. It does not have to be proficiency assessment. In fact, most of summative assessments are known as being achievement, norm-referenced and fixed-point assessment (Council of Europe, 2001, p. 186).

Summative assessment seems to be the most relevant to the Common Reference Levels. Yet, feedback provided from a summative assessment can serve as formative and diagnostic, which is apparent in the DIALANG Project (Council of Europe, 2001, p. 186).

Direct Assessment / Indirect Assessment. As the name suggests, direct assessment means assessing learners directly. To exemplify, the assessor observes a group of learners by comparing their performance with a grid and choosing the most appropriate categories for the group members on the grid, and finally gives a grade, which is what is done while the instructors offering writing course are grading the writing tasks of the students at the SFL in question. Comparing criteria to the match can be a way of assessing linguistic range and control directly. Receptive activities cannot be assessed directly. Herein, direct assessment is restricted to listening, writing and speaking interaction. An example for a typical direct test can be interviews (Council of Europe, 2001, pp. 186-187).

As to indirect assessment, it assesses skills by means of a paper test. Interpreting the responses of the learners to test questions can be a way of assessing linguistic range and control indirectly. As stated above, listening, writing

and speaking interaction are assessed directly. Yet, reading is merely assessed indirectly as it requires learners to provide evidence of their comprehension by answering questions, ticking or crossing boxes and completing sentences. An example of a typical indirect test can be a close test (Council of Europe, 2001, p. 187).

Performance Assessment / Knowledge Assessment. In performance assessment, the learner is supposed to contribute to the process by supplying a sample of their language use either in speech or written form, which makes performance assessment one of the assessment types used with the tasks carried out in the study. But in knowledge assessment, the learner is supposed to contribute to the process by answering questions on different issues, which are considered as evidence on their linguistic knowledge and control (Council of Europe, 2001, p. 187). One of the most important attributes of performance assessment is that its engaging learners with real learning activities like problem-solving activities and written communicative skills with the aim of modeling real-life does not deflect instruction (Gipps, 2012).

According to Palm (2008), performance assessment can be described by two definite categories: response-centred performance assessment and simulation-centred performance assessment. In the response-centred model, learner responses are related to written assessment tasks, and assessments in this type are commonly related to traditional assessment forms such as word completion tests and online multiple choice answers as a part of e-assessment. Simulation-centred performance assessment, on the other hand, involves observation of learner performance with professional equipment regardless of paper tests. This can serve authentic assessment activities as their assignments are related to real-life task measurement (Palm, 2008).

The distinction between performance and knowledge assessment is like the one direct and indirect tests have. As known, competences cannot be tested directly. Performances help generalise proficiency, and proficiency is the version of competence put to use. Considering all these, it can be said that all tests are able to assess performance only (Council of Europe, 2001, p. 187).

But for all that, an interview necessitates 'performance' more than filling in the blanks does, and filling in the blanks necessitates 'performance' more than multiple choice. That means 'performance' is there to refer to language production, not in a limited sense as it is used in 'performance tests'. The word is used for a study-related situation. In some tests, it is possible to see a balance between the performance assessment and the knowledge assessment, but for some of them, it is not (Council of Europe, 2001, p. 187).

Subjective Assessment / Objective Assessment. Assessors judge the quality of the performance in subjective assessment. And in objective assessment, there seems to be nothing subjective, like a multiple choice test the items of which have just one correct answer. Grades in direct performance assessment are decided upon a judgement. By this way, related factors are taken into account and referred to a criteria with the aim of deciding how well learners carry out the process, which also states that the decision is subjective. One advantage of this type of assessment is that language and communication are very complicated, and they are of great importance than some of their parts (Council of Europe, 2001, p. 188).

Subjectivity/objectivity issue is notably confusing. When the marker uses a key to accept or reject a response to a question and correct responses are counted for the exact result, an objective test is often called "an indirect test". A few test types such as c-tests and multiple choice have this process under control by having just one correct answer for each question, and to get rid of marker error, machine marking is additionally made use of (Council of Europe, 2001, p. 188).

All assessment is supposed to be objective. While selecting contents and performance types, personal judgements should be avoided, because third parties use test results very often to come to a conclusion about learners' future. The first step to reduce the subjectivity in the assessment is to establish a common framework of reference since the Framework tries to create a base for "the *specification for the content"* and "the *specific defined criteria* " in terms of direct tests (Council of Europe, 2001, p. 188).

Rating on a Scale / Rating on a Checklist. Rating on a scale means that the learner is decided to be at a specific level on a scale which has been made up of levels and bands. Placing the learner, who has been rated, on bands is important:

the emphasis is vertical: scale descriptors should have different levels / bands clear. Different categories may require several scales, which may be introduced on the same page or on different ones. Definition for each level, the top, bottom or middle is possible (Council of Europe, 2001, p. 189). Therefore, rating on a scale is one of the assessment types used at the SFL while grading the writing tasks of the students.

Rating on a checklist, on the other hand, is an alternative and it means that the learner is graded with regard to a list of points which are related to a specific level. The emphasis is horizontal: it is important to show how much of the module the learner achieved. The checklist can be in a questionnaire form or presented as a wheel. The answers may be Yes / No or more differentiated as the descriptors involve distinct and criterion statements (Council of Europe, 2001, p. 189).

Impression / Guided Judgement. Impression is simply based on learner's performance in class and assessor's subjective judgement on it regardless of any specific criteria with regad to any specific assessment. In-class performances and homeworks are good examples for impression as teachers assess learners based on these. Many subjective rating forms, particularly the ones in continuous assessment, include rating an impression based on reflection focused by the observation of the learner for some time. Many schools use this basis (Council of Europe, 2001, p. 189).

Guided judgement, on the other hand, is based on specific criteria, which helps reduce assessor's subjectivity by completing impression. It is used when impression is changed and guided into a judgement through assessment. This assessment approach suggests "(a) an assessment activity with some form of procedure, and/or (b) a set of defined criteria which distinguish between different scores or grades, and (c) some form of standardisation training" (Council of Europe, 2001), which is exactly the type of assessment the instructors offering writing course make use of while grading tasks in the study. Consistency of the judgements can be totally gotten better if a common framework of reference is brought into being for the assessors (Council of Europe, 2001, p. 189).

Holistic / Analytic. Holistic assessment is about having a universal synthetic judgement while analytic assessment means taking different aspects into account separately. It is the type of assessment used with the rubric in the study.

The distinction between holistic and analytic assessment can be made clear in two ways, and this much is certain that systems, from time to time, integrate an analytic approach with a holistic one:

(a) With regard to what is looked for; this is about what to assess as some approaches tend to assess global categorises like interaction, speaking for which they give one grade or score while analytic ones expect the assessor to have separate results for different parts of the performance in question. But still, there are other approaches requiring the assessor to mark down a universal impression, to analyse categories and finally to come up with a holistic judgement. Since separate categories of analytic approach give the assessor the chance to observe closely, it is advantageous. They also help create a metalanguage for feedback that learners need and between assessors for negotiation. One downside of it is that assessors are not able to separate categories from a holistic judgement with ease. Getting them to have more than five categories also makes them have "cognitive overload" (Council of Europe, 2001, p. 190).

(b) With regard to how a score or grade is attained; evaluating the results; some approches, in a holistic way, link observed performance with descriptors on a scale regardless of the scale's being holistic or analytic. There seems to be no arithmetic in those kinds of approaches. A single number or a 'telephone number' is used to report the results across categories. Some other analytical approaches necessitate to assign a mark for different points after which they require to add those points up in order to give a score to be turned into a grade (Council of Europe, 2001, p. 190).

Series Assessment / Category Assessment. Category assessment includes only one assessment task by which it is possible to judge performance with regard to the categories stated in an assessment grid.

Series assessment, on the other hand, includes a number of isolated assessment tasks like roleplays that are assessed with a basic holistic grade on a scale. Series assessment can be seen as a way to deal with the inclination in category assessment to put outcomes on one category, and to include others on another one. The emphasis at lower levels are mostly seen on task achievement with the purpose of completing a checklist about the learner's ability on an actual performance based on teacher/learner assessment. Tasks at higher levels are planned to have some specific aspects of proficiency in the learner's performance. A profile is used to report results (Council of Europe, 2001, p. 190).

Assessment by Others / Self-Assessment. Assessment by others, as the name suggests, means teachers or examiners do judgement part. Self-assessment, on the other hand, means learners judge their own proficiency.

Learners can make use of many assessment types stated above. According to researches, self-assessment - irrespective of 'high stakes' - can be useful as a powerful complement for teacher assessment and tests. There are two ways to increase the accuracy in self-assessment: (a) assessment is supposed to be in regard to explicit descriptors which define proficiency standards and (b) assessment is supposed to be connected to a specific experience, which can be even a test. If learners have the opportunity to receive training, the accuracy level gets higher (Council of Europe, 2001, p. 191). Additionally, self-assessment is believed to promote student autonomy by suggesting that it may have a disciplinary effect disempowering authorities (Tan, 2004).

Self and peer assessment cannot be underestimated in terms of their contributions to students' learning process. They have purposes which can be hardly accomplished in any other ways (Black et al. 2004, p. 12). These are primarily:

- helping to have a general understanding on assessment criteria and processes;
- giving learners the chance to monitor their learning process and progress;
- coming up with a reflective approach in order to enable meta-cognitive skills;
- and bringing responsibility and confidence in learners.

According to Black (2004), self-assessment is a difficult skill to acquire, so learners' undertaking peer assessment first can be a constructive way like a bridging skill, leading to self-assessment. When learners assess their peers' work, they try to improve their work as this process serves as a stimulus to this (Black et al. 2004).

Conclusion

The CEFR can be perceived as a redefiniton needed for language learning and teaching via plurilingual approach. It serves to enhance international communication, to promote mobility and to increase respect and tolerance for cultural diversity. Through scientific research, the CEFR provides a common and practical tool for having clear standards at the stages of language learning and language testing. With the aim of describing what a language learner is expected to do at each level of proficiency, communicative competence theories are drawn on. These make the CEFR a key reference source and a precious tool for education, and is of interest to those being involved in language learning, teaching and testing.

The CEFR described proficiency levels in a clear way and provides guidelines on how language education is supposed to be carried out. The language policy of the Coe is to have competent learners in communicating different languages, so the ultimate goal is "to achieve greater unity among its members" (Council of Europe, 2001, p. 2).

The focus of this thesis has been the A2 Level - Waystage. The objective at this level is to communicate with both the native and non-native speakers to exchange information on everyday issues. The learners at this level are supposed to perform certain transactions, give and get personal information, and lastly forge closer ties with people in terms of professional and social contacts.

With these thoughts in mind, the research presented in this thesis aims at analysing writing skill studies at the Waystage level and developing them by means of the data gathered from both the instructors offering writing course and the learners.

Chapter 3
Methodology

Introduction

The methodology that was used for this study is described in this chapter. It starts with the theoretical framework rationalizing the research design that was followed. After that, the description of the setting and participants are presented, and an explanation on the instruments used in the study is provided. Next, the data collection process is depicted thoroughly.

Theoretical Framework

In this study, a correlation research design was followed where relationship between variables is determined, and in the event of a relationship, predictions to the population in question are made (Simon & Goes, 2011). This correlation study was applied in order to determine if there is a significant difference between the gradings of the instructors offering writing course and the standardisation team. Just one experimental group participated in the study, and convenience sampling, which is a nonrandom sampling type, is used while selecting the group (Mackey & Gass, 2005).

The methodological approach in this study is the one serving both qualitative and quantitative data collection techniques, which is called a "mixed methods research design" (Christensen, Johnson & Turner, 2015). As noted by Johnson and Onwuegbuzie (2004), both qualitative and quantitative traditionalists appraise their own model as the best one for research. Yet, both have drawbacks. To illustrate, one of the aspects of quantitative methods, exploratory, is not found sufficient as it falls short of showing the reasons behind research studies. Brannen refers to this method as decontextualized and simplistic considering its generalizations (as cited in Dörnyei, 2007). On the other hand, many others consider qualitative methods as inapplicable, time-consuming and blurry (Dörnyei, 2007). Another drawback of the qualitative method is researcher bias. Mixed method research draws on the strengths of both qualitative and quantitative, and minimizes the partition between them (Johnson & Onwuegbuzie, 2004). The advantages of mixed methods research design are stated by Denscombe (2008) and Dörnyei (2007) as; reaching multiple audiences, reducing single methods' weaknesses, providing more accurate data and creating a better understanding on

the idea that is under investigation. Besides, it serves 'triangulation', which was stated as an influent way for research validity by Dörnyei (2007). Other than these, Johnson, Onwuegbuzie and Turner (2007) also listed a few more reasons such as reaching more comprehensive findings and internal consistency, providing deeper and better understanding as well as a better description, and last, getting more useful and meaningful answers.

Creswell (2012), Teddlie & Tashakkori (2009) and Creswell & Plano (2007) specified the distinctive characteristics and practices of the mixed method research design in their works. Creswell (2012) featured six mixed methods designs, and an embedded design, in which the main data source is supported by a secondary one, has been adopted. For the major purpose of the study, the gradings of the instructors offering writing course and the standardisation team were compared via SPSS so as to see if they conform with each other. For the secondary purpose, qualitative data (interview with the instructors offering writing course and comparing rubrics via comparative and thematic analyses) was used to extend more on the quantitative results. The reason why the secondary data was collected is because the major data source is intended to be provided with additional information.

Setting and Participants

The study was carried out at Karabuk University, School of Foreign Languages in Turkey. The data collection tools were administered to 93 preparatory class A2 level students and 4 instructors teaching writing to this focus group in writing classes. In the departments, the medium of instruction is English, so before students proceed to their departments, they are required to study English language in the SFL. Yet still, if students get at least 75 pts. from the exam carried out by the university itself or if they hand over a valid international English proficiency exam result such as TOEFL IBT or a national one such as YÖKDİL or YDS, they are held exempted from the preparatory school. If that is not the case, they are required to enroll in the preparatory school and study English for three semesters. If failed in the proficiency exam at the end of the academic year, students are considered as 'repeat students' and they study English in the

preparatory school for one more year. In the event of failure again, students are expelled from the university. If not, they proceed to their departments.

The preparatory school aims to graduate qualified individuals to be studying at various departments. It prepares students to be certified proficients in English language and to use their language knowledge in their departments properly. The program also helps them gain practical and cognitive skills, the skill to work responsibly and essential competencies on communication, learning and socialising.

A nonrandom sampling method, convenience sampling, was used to select the participants. Convenience sampling was defined as a group of (conveniently) available people by Fraenkel, Hyun, and Wallen (2012), and the current study was carried out with the students that the volunteer instructors offering writing course were teaching. The participants are A2 level students in 2018-2019 academic year. All of them signed the consent form before the commencement of the study. The students study with 4 non-native instructors for 26 hours a week, and two hours of classes are for writing. 24 of the students are female and 69 of them are male. Their ages range from 17 to 20. 88 of them come from different cities of Turkey and 5 of them are from Syria. All of the students could fill in the A2 level questionnaire, but some of them could not take some writing tasks. There are missing scores in the writing tasks because of absenteeism, and they were not allowed to take the ones they had missed.

As for the interview participants, a non-probability sample, purposive sampling strategy was made use of. As Patton (2002) stated, why purposive sampling is needed is because of the importance placed on in-depth understanding. The participants were selected with the aim of getting a better understanding of the phenomenon in detail, and among the various purposive sampling strategies, the convenience sampling also known as availability sampling (Patton, 2002), was employed to collect information from the population to be interviewed.

There are 65 non-native (Turkish) instructors of English at the university in question, and their teaching experiences range from 1 to 14 years. There are four offices responsible for meeting various needs of the SFL. They include: the

material office, the standardisation office, the testing office and the level coordinating office.

As regards the details about the participant instructors offering writing course, they teach approximately 24 hours a week, and have 2 hours as writing lesson. 2 of the participants are female and the other 2 are male. They all graduated from English Language Teaching departments, and have been continuing their master's degrees at different universities. Two of them have been working at Karabuk University for 5 years, and the other two for 8 years. In the table below, the necessary information about the instructors offering writing course is presented:

Table 12

The Instructors

	Age	Years of Experience	Master's Degree
1. Male	28	5	✓
2. Male	35	12	✓
1. Female	29	7	✓
2. Female	32	9	✓

Data Collection

Right after Hacettepe University Ethics Commission granted the necessary approval stating that the study was in conformity with the ethical principles, data collection process initiated. Participants for the study were chosen through convenience sampling. As the researcher was a member of the standardisation team in the School of Foreign Languages at Karabuk University, she thought it would be more useful to work on the CEFR practices and standardisation studies. After the consent of SFL administration, the details were discussed with the standardisation team members and the volunteer instructors offering writing course. As they all were striving for improvement in the school concerning the CEFR practices, they got thrilled by the idea.

The data collection process consists of three phases (Table 15). The students and the volunteer instructors offering writing course were informed about the CEFR, the study and its purpose briefly. The information session was carried out by the researcher herself in four different classes. The students and the instructors were explained that they were free to drop out the study anytime they feel uncomfortable. The volunteer instructors were also informed that they would be asked to be interviewed after they agreed to take part in the study. All participants were assured that the data gathered from them would be used for nothing but scientific purposes, cause no negative effects on them, and no other institutions or people would know about them. After the information session, the participant students and instructors were given an official consent form in which their rights were expressed clearly.

Table 13

Data Collection Process

Data Collection Process	
	Introduction of the study
Phase 1	Consent forms
	Can-Do Statements questionnaire
	In-class
Phase 2	Writing tasks
	After-class
	Instructors' grading tasks
Phase 3	Rubrics
	Interview

Next lesson, "can do" statements questionnaire was distributed to the participant students. They were explained how important it was for them to be honest while answering questions, and the results of this questionnaire would cause no consequences for them. That instrument was conducted in order to determine participant students' perceptions of their own English language proficiency. All the students could fill in that form as the instructors guided them through the process. The time allocated to complete it was enough; the students

delivered their papers when they were done with them. Even if some students were absent, they had the chance to fill in form the following day. A total number of 93 students were involved in the process.

During the term, the students had five in-class writing studies which were not graded by the instructors, but given immediate feedback on. The purpose of these studies was to have students see their weaknesses and strengths in writing. Other than these studies, they had two different writing tasks that were graded by the instructors offering writing course and taken as exams. The underlying premise here is that students become involved in writing studies first in order to get fully prepared for the writing exams; they have the chance to ask for information on their mistakes, how to correct those mistakes and get feedback again after their attempt to correct them.

After writing task exams, the instructors offering writing course graded the papers in accordance with the rubric they had been using. The instructors were given 4 days to finish grading papers, after which they were required to hand their writing paper packs over to the standardisation team for the second grading. After another 4 days, the standardisation team finalized gradings and according to the procedure, if the grading of a paper by any instructor and the standardisation team members were four points apart, the team would arrange a meeting with the instructor and discuss the situation trying to find a common ground. If the gap between gradings were less than four points, then the grading by the instructor would remain valid. The same applied to all gradings by all instructors offering writing course.

In this study, there were some missing papers due to absenteeism. Namely, for each writing task during data collection, the number of students taking part in the process was different since some students did not attend the lessons that day. The absentees were not many, yet the results were defined in percentages in order to have a clear picture irrespective of the changing number in students' papers.

After the results of the gradings by the instructors offering writing course and the standardisation team had obtained, the rubric that the instructors in Karabuk University had been using for a few years was analyzed and compared

with the one based on the CEFR. The standardisation team members had a session with the researcher about the rubric explaining how it had been formed by the team, what kinds of sources they had made use of building it, how effective they saw it had been considering the needs of the students and how much it had served the purpose. The writing task assessment rubric used by the instructors can be found in Appendix F.

The participants were chosen by purposive sampling as stated in the setting and participants part. At the end of the term, the instructors who offered writing course to the focus group were interviewed about their perceptions on the CEFR descriptors and how much they know about the CEFR. Each interview took approximately 10-15 minutes. The interview session had to be carried out in school as the instructors offering writing course were all present there and the meeting had to be scheduled to take place in just one day. The interviewees were taken to the meeting room, which had been arranged by the school administration, one by one out of class hours. They were ensured that what was uttered by them in the interview session would not be shared with anyone, and their identities would remain unrevealed. All interviewees were observed to be willing and sincere while being interviewed. The questions in the interview can be viewed in Appendix G.

Instruments

The instruments used in this study are (1) a "can do" statements questionnaire to determine the level of the focus group (see Appendix C), (2) writing tasks graded by the instructors who offered writing course to the focus group (see the example on Appendix D), (3) the writing task assessment rubric in accordance with the CEFR descriptors to assess A2 Level writing tasks (see Appendix E), (4) the writing task assessment rubric used by the instructors to assess A2 Level writing tasks (see Appendix F) and (5) interview with the instructors offering writing course (see Appendix G).

Can-Do Statements Questionnaire. "Can do" statements questionnaire developed by the ELP to determine the levels of learners, and has seven questions with a four-point Likert type scale. It was used to determine the focus group's level in this study based on their own perceptions. There are seven items in the questionnaire and the students were expected to answer them by choosing a number from 1 to 4, 1 as "rarely", 2 as "sometimes", 3 as "often" and 4 as "always". By this way, students were involved in the process and they had the chance to assess themselves. Three outside experts working in the EFL field verified the validity of the instrument.

The items in the questionnaire aim to assess the level of students from their own point of view since it is important to know how the students perceive their progress and ability in language learning. The items are the expressions like "I can create notes about where I am, or where and when to meet somebody" (BEDAF, 2012). and they give the students a chance to evaluate themselves on what they can do in the target language.

Table 14

Layout of the Questionnaire

Question	The Point Surveyed	Number of İtems
1	Creating notes	1 Likert-type item a four-point scale
2	Writing a short letter	1 Likert-type item a four-point scale
3	Describing family, hobbies, etc.	1 Likert-type item a four-point scale
4	Describing an event/activity	1 Likert-type item a four-point scale
5	Writing short stories via pictures	1 Likert-type item a four-point scale
6	Writing short texts	1 Likert-type item a four-point scale
7	Filling in forms	1 Likert-type item a four-point scale

As seen above, the questions were designed with the purpose of determining how learners perceive their writing skill; how well they believe they do in certain areas of writing in the target language and to what extent they believe

they do so. The questions start with one of the basic ones, asking students about their ability to create notes in the target language, and they proceed with more complex ones; writing a short letter, describing hobbies, family and an activity/event, writing short stories by using pictures, writing short texts, and last, filling in forms.

Writing Tasks. Writing has to be a major concern and interest for researchers, students and teachers as it is one of the main pillars of language learning (Sim, 2010), so one of the most important parts of the data collection process in the study is the writing tasks the focus group students dealt with. Each of these tasks lasted for one hour and the students took them like an exam.The instructors offering writing course to the focus group graded those papers, and the standardisation team regraded them, as well. Then, it became possible to see the difference between the grades given by the instructors and the standardisation team. As the standardisation team has been commissioned by the school administration and got special training to carry on the practices at school based on the CEFR principles, it is of vital importance to get the grades regraded by them. The writing task topics were as follows:

Table 15

Writing Tasks

	Writing Task 1	Writing Task 2
Type of Text	Narrative	Descriptive
Min. Number of Words	80	80
Topic	A special occasion	An ideal university
CEFR Domain	Personal	Educational
CEFR Descriptor	Correspondence	Correspondence
Envisaged CEFR Level	A2	A2

There are three semesters in one academic year at Karabuk University, and in each term students are required to take writing task-exams, each lasting for one

hour. The first two semesters, the students are asked to write about tasks such as information request (letter), invitation to a party, school days, their diet, their favourite photograph, holidays (letter writing). Last term, they are asked to write about tasks such as description of their hometown, their last or best holiday, description of a family member …etc. in paragraph form. As seen in the table above, the students are asked to write a narrative paragraph about a special occasion and a descriptive one about an ideal university in the last term. The number of word limit is 80 for both, and the CEFR domain for the narrative paragraph is personal whereas it is educational for the descriptive paragraph.

The Writing Task Assessment Rubrics. Both structural and grammatical parts in writing are illustrated differently at different levels of proficiency by means of a practical performance criteria known as rubrics (Stiggins, 2001). For this purpose, a writing task assessment rubric was introduced in accordance with the CEFR descriptors. The differences between that one and the one used by the instructors since 2014 were discussed in the data analysis part. The CEFR-based rubric was carefully designed to meet the needs in the writing lessons at A2 level. If there is no objective criteria determined prior to grading, instructors might have to trust their own instincts. To avoid this problem, a CEFR-based rubric is needed in assessing writing in order to achieve objectivity aimed when writing tasks are graded.

Interview with the Instructors. As Seidman (2006) stated interviewing means finding out the experiences people go through and understanding the meaning those people make of it. Perakyla and Ruusuvuori (2011), on the other hand, asserted that interviews give the researchers the chance to get into the 'reality areas' such as people's personal attitudes and experiences which would otherwise stay beyond reach. For this study, other than that, having both quantitative and qualitative data enables the researcher have better results supporting one another. With this thought in mind, a semi-structured interview was prepared so as to gather information on the focus group writing lesson instructors' perceptions on the CEFR and the process. The interview, as seen in the Appendix G, involves 5 questions and they were centered on (1) what the instructors offering writing course know about the CEFR, (2) instructors' viewpoints on the benefits of assessing writing in accordance with the CEFR descriptors, (3) instructors'

viewpoints on the contribution of the CEFR to teacher development, (4) whether they think that the rubric they have been using is based on the CEFR descriptors, (5) how they think the spotcheck done by the standardisation team help them and whether their gradings conform with the standardisation team's. Experienced instructors were chosen as this study needed experienced ones to comment on CEFR practices. Their years of experience range from 5 to 12 years, which is satisfactory as Kelchtermans & Vandenburghe (1994) stated that teachers with at least 5 years of practice are assumed experienced teacher. The interview with each instructor took approximately 10-15 minutes and the answers by the instructors offering writing course to the focus group were transcribed. The way the interview was carried out was described in the data collection section.

The instruments above were intended to achieve the following objectives:

- To gather information about the students' perceptions on their ability in the target language
- To analyze the similarities and the differences between the rubrics the instructors have been using and the one based on the CEFR descriptors
- To investigate the instructors' attitudes and perceptions on the process and the CEFR practices
- To analyze the similarities and the differences between the instructors and the standardization team on grading the writing tasks of the focus group students.

Data Analysis

Introduction. The purpose of this study is to determine how effectively and to what extent the CEFR descriptors are used while assessing the writing tasks of the A2 Level students at Karabuk University, School of Foreign Languages. This chapter shows the results of the data collection procedure administered in the 2018-2019 Academic Year. The responses of "can-do" statements questionnaire were analyzed via SPSS 22. The answers of 93 students for each statement were analyzed, and mean scores and standard deviations of them were calculated. The same way was used for the writing task grades: the grades by the instructors and the standardisation team were compared and mean scores and standard deviations of them were calculated.

A comparative and thematic analysis was used to indicate the differences and similarities between the rubric that has been used by the instructors and the one based on the CEFR. The data in hand were identified, first. Then, the data were analyzed, and as the last step, the patterns were reported within the data (Braun & Clarke, 2006).

As for the semi-structured interview, open-ended questions were chosen so as to "allow the respondents opportunities to develop their responses in ways which the interviewer might not have foreseen" (Campbell, McNamara & Gilror, 2004). Interview is believed to allow for the exploration of some issues as they arose (Kvale, 1996), and qualitative content analysis was used for it through which the data is quantified. In this analysis, which is based on the latent content analysis of Dörnyei (2007), qualitative categories are not determined in advance, but attained from the analysed data inductively. Correspondingly, Perakyla and Ruusuvuori (2011) emphasized that there is not a predetermined protocol to create main themes; the raw data are read repeatedly. The transcriptions in this study were reread once and again accordingly, and in order to define categories and key themes, the coding of the transcribed data was performed. According to Miles and Huberman (1994), codes are labels used for assigning meaning units to the inferential or descriptive information gathered from a study… connected to words, chunks, sentences, phrases or paragraphs. As regards the benefits of coding, Maxwell (2005) stated that it enables the researchers to compare the deduction of theoretical concepts and the data easily, and provides strong definitions. The details about the data analysis of the instruments are as follows.

Can-Do Statements Questionnaire. First, a Can-Do Statements questionnaire was used to get a clear profile of the students' perceptions on their level of writing proficiency. The main purpose was to gather information about their perceptions on their ability in the target language and to find out if they really see themselves at A2 Level as categorized by the school administration at the beginning of the term. As mentioned earlier, 93 students filled in this questionnaire. The questionnaire was prepared by the ELP to determine the levels of learners, and has seven questions with a four-point Likert type scale. The descriptive analysis of the results were presented in Table 16.

Table 13

ELP Can-Do Statements Questionnaire

	Number of Ss.	Mean	Std. Deviation
Q1	93	3,5376	,71565
Q2	93	3,4301	,78571
Q3	93	3,5591	,80030
Q4	93	3,3656	,81807
Q5	93	3,0968	,84782
Q6	93	3,4946	,70130
Q7	93	3,4301	,68203

It is seen that the students in the focus group have positive opinions on their language ability. Considering the mean of each item above, it can be said that students also perceive themselves as A2 Level students. If we look at the distribution of the students' answers , we have a table like the one below:

Table 14

The Distribution of the Students' Answers

		N	Mean	SD.
	4	60		
Q1: I can create notes about where I am, or where and when to meet somebody.	3	25	3,5376	,71565
	2	2		
	1	1		
	4	55	3,4301	,78571
Q2: I can write a short letter to express my thanks, to apologise, to send greetings.	3	25		
	2	11		

	1	1		
	4	66		
Q3: I can describe my family, myself, my hobbies, my school or my job in short texts.	3	17	3,5591	,80030
	2	6		
	1	1		
	4	52		
Q4: I can describe an event or an activity such as a celebration or a meeting in simple sentences.	3	25	3,3656	,81807
	2	14		
	1	2		
	4	36		
Q5: I can write short stories using pictures.	3	32	3,0968	,84782
	2	23		
	1	2		
	4	57		
Q6: I can write short texts with intro., development and conclusion parts, using connectors.	3	25	3,4946	,70130
	2	11		
	1	-		
	4	50		
Q7: I can fill in a form about my educational background, my job, my fields of interest and skills.	3	33	3,4301	,68203
	2	10		
	1	-		

As seen above, the students in the focus group find themselves more confident and able while writing about their personal life and social environment. If pictures are involved in the writing activity, some students abstain. Yet, all sudents believe that they are good at writing short letters, stories, texts, creating notes,

describing events and filling in forms, which makes them target language learners at A2 Level.

Writing Tasks. As stated above in the "instruments" part, the students at the SFL had two writing tasks performed like exams taking one class hour. The first one was a narrative, and the second one a descriptive paragraph. While carrying out those tasks, the instructors let the students use their dictionaries, and after each task, the instructors graded the papers. As the last step, the standardisation team regraded the papers to make sure they had been graded in accordance with the writing assessment rubric prepared by the office and the grades conform with the standardisation teams'. As for the assessment, continuous, formative, direct, performance, rating on a scale, guided judgement and analytic assessment types are adopted at the SFL. In the interview, the instructors stated that they believe their grades conform with the standardisation office's, and the results show that they do. The descriptive analyses of the grades given by the instructors and the standardisation team are detailed in the findings part.

The Writing Task Assessment Rubrics. The rubrics help improve communication about the value of the assessment and the grading of learners work (Rasheed, Aslam & Sarwar, 2010). The writing task assessment rubric that has been used by the instructors of Karabuk University for this purpose and the one based on the CEFR descriptors have a few points in common while having a few differences as well.

As stated in the introduction part, a comparative and thematic analysis was used to indicate the differences and similarities between the rubric that has been used by the instructors and the one based on the CEFR descriptors. The rubric used by the instructors has five categories to assess different points in the target language, which are "content, organization, vocabulary, grammar and mechanics." The CEFR-based one has exactly the same categories with exactly the same points for each category (content: 5 pts, organization: 5 pts, vocabulary: 4 pts, grammar: 4 pts, machanics: 2 pts, and 20 pts. at total). The data on the differences and similarities between these categories were identified, analyzed and reported as shown in the data analysis part.

Interview with the Instructors. As stated in the 'instruments' part, an interview was carried out with the instructors offering writing course to the focus group students. The instructors were asked 5 questions, and the questions were about the CEFR descriptors and the perceptions of the instructors offering writing course on the CEFR practices in general.

The core of the data analysis is constituted with the inductive analysis of the interview between the instructors and the researcher. The interview was analyzed in the light of what the instructors offering writing course mentioned about the CEFR. The verbatim transcriptions of the interview were worked on through content analysis to categorise the collected data and thematic analysis which is based on constant comparison. Namely, the transcribed data were initially coded by adopting content analysis. Those codes were given meaning. Thus, an instructor's statement: "The CEFR helps us improve our teaching" was coded as *"improving teaching"* while another one "The CEFR helps teachers change the idea of teaching" was coded as *"changing the idea of teaching".* After coding, these codes were grouped into categories. The two codes mentioned above were categorized as *"teacher development".* After categorising, they were grouped into themes as *"assessment"* and "*professional development"* as described in the coding method by Strauss & Corbin (1990).

In order to provide reliability, first, the interview's transcribed data were read through by the researcher and the peer coder separately, then the data were coded independently. Following this, they worked on the codes together and tried to reach an agreement on them. The procedure was followed for the rest after they agreed on the codes to a large extent. Next, the codes were compared. What 'agreement on the codes' means in this study is that both coders appoint a very similar or the same word to a text segment. Then, some categories and themes were identified, after which the coders matched the codes with the right themes and categories separately. The calculation of the frequency of recurring codes was carried out, and agreement between the coders was determined. As Miles and Huberman (1994) stated, check-coding is a great way to check reliability. Intercoder reliability, as Lombard, Snyder-Duch and Bracken (2002) asserted, shows to what extent independent coders reach an agreement on the same

codes. Hence, a 90% coding agreement, which was recommended by Miles and Huberman (1994), was aimed in this study.

Chapter 4
Findings

This chapter merely focuses on the data obtained by means of the data collection tools aforementioned and the research questions restated below. Each analysis is described through the qualitative or quantitative data analysis methods. With the data gathered, a better understanding about assessing writing in accordance with the CEFR descriptors will be provided. It is also hoped that the findings in this thesis will help develop a better assessment system with regard to writing tasks.

The research questions are as follows:

1. Are the writing tasks of the A2 level prep-class students assessed in accordance with the CEFR descriptors?
   a) Are the rubrics that the instructors offering writing course use based on the CEFR descriptors?
2. To what extent do the instructors' gradings conform with the CEFR standardisation team's gradings?
3. What do the instructors offering writing course know about the CEFR?
4. What are the instructors' viewpoints on assessing writing in accordance with the CEFR descriptors?
5. What are the instructors' viewpoints on the contribution of the CEFR to teacher development?

With the help of the "Can-Do" Statements questionnaire, 93 students belonging to the focus group were determined as A2 Level university students and accepted as the subjects of the study. As noted earlier, five research questions were formulated. Quantitative data were analyzed via SPSS 22, and qualitative ones were transcribed for the sake of clarity. Then, they are compared and coded. Lastly, the codes were grouped into categories and then themes to have a better understanding as Silverman (2000) suggested, imposing prior categories of analysis and forming such premature categories were avoided.

Research Questions 1: "Are the writing tasks of the A2 level prep-class students assessed in accordance with the CEFR descriptors?"

    a) "Are the rubrics that the instructors offering writing course use based on the CEFR descriptors?"

This question and its sub-question are explained together as they are connected and a part of each other. The rubric that the instructors used while grading A2 Level writing tasks (Appendix F) was analyzed and it was compared to the one based on the CEFR descriptors by means of comparative and thematic analyses (Appendix E). Tables below indicate the differences and similarities between two rubrics.

Table 15

The Parts of the Rubric Used by the Instructors

	The rubric used by the instructors					
	0 pt.	1 pt.	2 pts.	3 pts.	4 pts.	5 pts.
Content	No relevance	Considerably irrelevant	Inadequate coverage Inappropriate ideas	Adequate coverage Appropriate ideas	Good coverage Clear ideas	Very good coverage Fully developed passage
Organization	Completely disorganized	Inappropriate Format	Inadequately organized	Adequately organized	Generally well-organized	Well-organized
Vocabulary	Very limited range of vocabulary	Translation-based mistakes	Some use of varied word choice	Adequate range of vocabulary	A wide range of vocabulary	
Grammar	Lack of language	"Rarely clear language Defects in complex constructions	Several problems with grammatical structures	No major difficulties	A wide range of structures Few mistakes	
Mechanics	No control	Occasional mistakes	Masters conventions			

Table 16

The Parts of the Rubric Based on the CEFR

	The rubric based on the CEFR					
	0 pt.	1 pt.	2 pts.	3 pts.	4 pts.	5 pts.
Content	No relevance	Few content points mentioned	Some content points mentioned Some valid ideas	Several content points mentioned Reasonable ideas	Most content points mentioned Valid ideas	All content points mentioned Wide range of valid ideas
Organization	Completely disorganized	Rare use of simple linking devices	Inadequately organized ideas	Adequately organized valid ideas Some simple linking devices	Clearly well-organized valid ideas Appropriate linking devices	Effectively organized valid ideas A wide range of appropriate linking devices
Vocabulary	Very poor / insufficient vocabulary	Mistakes in using some simple, short everyday chunks	Mistakes in linking groups of words with simple connectors	Few mistakes in linking groups of words with simple connectors	Hardly any mistakes in linking groups of words with simple connectors	
Grammar	Serious lack of some simple language structures	Frequent mistakes in the use of simple structures	Mistakes in forming simple sentences Occasional mistakes; mostly interference from mother tongue	Few mistakes in forming simple sentences and phrases	Simple phrases and sentences linked well with simple connectors Systematically basic mistakes	
Mechanics	No control	Occasional mistakes	Very few mistakes			

To have a better understanding on the issue, two rubrics were compared and the table below shows the similarities and the differences between the rubrics. (To simplify the expressions, the rubric based on the CEFR is labelled as "A", the one used by the instructors is labelled as "B");

Table 17

Similarities and Differences between the Rubrics.

Parts	Similarities	Differences
Content	Wording is similar in some parts.	While refering to sub-parts, A uses *quantifiers*, B uses *adjectives,* and has more details.
Organisation	Wording is similar in some parts.	B has more details than A. B has vague expressions, A uses the word "simple" in each sub-part.
Vocabulary	Wording is similar in some parts.	A emphasizes "simple" word usage, B does not.
Grammar	Wording is similar in some parts.	A emphasizes "simple" structure usage, and is specific to A2 Level. B emphasizes complex structure usage, and is more general.
Mechanics	Almost the same	
Others	Both have 5 parts. Points for each part are the same.	

As it is seen, these two rubrics have a few things in common as well as a few differences. The details are as follows:

- In the content part, in the one used by the instructors, expressions like:" No relevance, little or no development of…, considerably," match up to the expressions like:" Hardly any or no relevance, few points mentioned" in the CEFR-based one. Yet, in the rubric based on the CEFR, the expressions are more of quantifiers (some, few, several) whereas they are qualitative in

the first one such as" irrelevant, appropriate, and adequate" as in the example below:

In the "content part" in the CEFR-based rubric, "2 points" accounts for "Some content points mentioned, and some valid but insufficient ideas in completing the task" while in the rubric used by the instructors it accounts for "Inadequate coverage of the topic, and addresses the topic using inappropriate and/or insufficient ideas".

- In the organization part, the expressions in the rubric used by the instructors are qualitative as well. The expressions seem to be more general whereas the ones in the CEFR-based rubric seem to fit well with the A2 Level CEFR descriptors, and "simple linking devices" are considered important while grading. To illustrate;

In the "organization part" in the CEFR-based rubric, "3 points" accounts for "Adequately organized valid ideas, mostly coherent and fluent, and some simple linking devices" while in the rubric used by the instructors it accounts for "Adequately organized (a weak topic sentence, unbalanced development of supporting ideas), and mostly coherent and fluent passage; ideas flow meaningfully and logically, but may contain some redundancy and some unclear connections."

- In the vocabulary and grammar parts, both show similarities. Yet, the one used by the instructors seems more general as it is used for all levels. In the one based on the CEFR, "forming simple structures" is considered important for A2 Level and the points are determined according to the extent that learners use those structures correctly. To illustrate;

In the "vocabulary part" in the CEFR-based rubric, "1 point" accounts for "Narrow, inadequate range of words, translation-based errors, and mistakes in using some simple, short everyday chunks" while in the rubric used by the instructors it accounts for "Uses limited range of words, most of which are inaccurate or irrelevant, and translation-based errors."

- The biggest similarity is on the mechanics part. Both rubrics have the same expressions with a slight difference in the last point: one uses "mastery in mechanics" while the other one based on the CEFR uses "very few mistakes" as for A2 Level learners it is considered natural to have those kinds of mistakes.

In light of this information, it can be said that the writing tasks of the A2 level prep-class students are intended to be assessed in accordance with the CEFR descriptors, and the rubric the instructors have been using was tried to be designed according to the CEFR descriptors, but there are a few differences. In the fourth question in the interview, the instructors offering writing course were asked if they believe the rubric they have been using is based on the CEFR descriptors. They said that they know it is based on the CEFR descriptors. Yet, the biggest and the most important difference is that even if the writing tasks are intended to be assessed in accordance with the CEFR descriptors and the rubric that the instructors have been using was designed according to the CEFR descriptors, the rubric in question is a general one. Namely, it is not specific for A2 level, but for all levels, so the features considered important in assessing writing for A2 level are needed to be looked from a broader perspective. To illustrate; for A2 Level, the students are required to form simple sentences with simple linking words, and systematic mistakes are considered natural as shown in the rubric based on the CEFR. Yet, in the one that is used by the instructors, simple forms are not mentioned, and even the benchmark "complex constructions" is included in the rubric.

Considered thoroughly, it can be said that the rubric used by the instructors is more general as it is used for all levels, and it could have been designed more appropriate regarding the forms that are specific to the A2 Level. The CEFR-based one was prepared just for A2 Level, which makes it easier to use while grading the writing tasks of the A2 Level students.

Research Question 2: "To what extent do the instructors' gradings conform with the CEFR standardisation team's gradings?"

In the interview, all instructors offering writing course stated that they believe their gradings conform with the standardisation office's. The descriptive analyses of the grades given by the instructors and the standardisation team are as follows:

Table 18

Descriptive Analysis of Task 1

	Number of Ss.	Mean	Std. Deviation	P
Instructor	91	16,82	2,4	
				0,097
Stnd. Team	91	16,20	2,5	

The first task carried out at A2 Level classes was about writing a narrative paragraph with at least 80 words. 91 students took the writing task exam, the other two were absent on the day of the exam. As seen in the table above, the p-value (Calculated Probability) is 0,097, which is bigger than 0,05, and this gives us the chance to conclude that no statistically significant difference was identified between the grades by the instructors and the standardisation team. Namely, the gradings of the instructors do not vary too much more than the ones by the standardisation team.

Table 19

Descriptive Analysis of Task 2

	Number of Ss.	Mean	Std. Deviation	P
Instructor	93	15,92	2,77	
				0,13
Stnd. Team	93	15,29	2,91	

The second task was about writing a descriptive paragraph with at least 80 words. As shown in the table, the analysis of independent t-test reveals that the difference between the grades by the instructors and the standardisation team is not significant, just like the first task. The variability in the grades by both groups is not significantly different (p: .13 / p: .05).

Taking the results into account, it is obvious that the difference between the grades by the instructors and the ones by the standardisation office is not significant. As it is seen in the descriptive analysis tables (see Tables 21 & 22), mean score for instructors for the first task ($\bar{x} = 16,82$) is slightly higher than that of standardisation team ($\bar{x} = 16,20$). For the second task, mean score for instructors ($\bar{x} = 15,92$) is again slightly higher than that of standardisation team ($\bar{x} = 15,29$), so this indicates there is not a meaningful difference between the mean scores of these two groups.

As for the opinions of the instructors about the standardisation team and their spotcheck system for the writing task grades, they mentioned the key terms below in oder to indicate the way the standardisation team helps them with.

Table 20

A Summary of the Findings Concerning the Respondents' Viewpoints on the Spotcheck Done by the Standardisation Team.

Codes	Number of Respondent	Percentage
Gaining experience	2	50%
Objectivity	4	100%
Criteria	2	50%

The instructors indicated that they believe the spotcheck done by the standardisation team help them gain experience, and they feel they are more objective while assessing writing tasks with the help of criteria the team provides. Their assertions support the results obtained via SPSS program.

In these two research questions, we focused on the results gathered by means of writing task assessments rubrics and writing task scores, and they were analysed. Firstly, the rubric used by the instructors was compared with the one based on the CEFR descriptors. That was important since the rubric was needed to be analysed in order to get better results while comparing writing task grades by the instructors and the standardisation team. Secondly, to find out if the gradings of the instructors offering writing course conform with the standardisation team's, the gradings of both group were compared via SPSS program and the results were presented.

Another important data source was the interview with the instructors. Analysing exam papers and rubrics was one part of the study, but the instructors' perceptions about the CEFR descriptors and practices were equally important. By this way, the quantitative data were combined with the qualitative one.

Research Question 3: "What do the instructors offering writing course know about the CEFR?"

An interview was held with the instructors offering writing course to the focus group students, and interviewees were chosen through a kind of non-probability sample, purposive sampling method. As explained in detail in the setting and participants part, convenience sampling was noted to be appropriate to specify the population. There were 5 open-ended questions in the semi-structured interview (see Appendix G). The analysis of this question merely revealed that the instructors offering writing course know about the basics of the CEFR and its practices. What they know about the CEFR is practical application of the CEFR practices rather than hearsay. At the beginning of the academic year, an in-service training of 4-days about the CEFR was organized by the administration, focusing on the practices and how to carry them out at the SFL in Karabuk University. The brochures and the school administration's citations in some other meetings also gave the instructors an idea of the CEFR as exemplified in the following statements by the instructors:

- "It is the classification of language proficiency in order to determine the level and readiness of the learners for the provided learning environment."

- "It is a framework that was created by the Council of Europe. It is used to describe what language learners can do at different stages. It has six stages from A1 to C2."

The answers above are from the instructors who have 5 and 7 years of teaching experience. Neither of them got training on the CEFR practices and descriptors before their in-service training. They also feel the need to learn more about it and carry out studies. They are self-confident young instructors, and they acquaint themselves with the CEFR practices day by day with the help of the school administration.

When asked the other two instructors, whose years of teaching experience are 9 and 12, they indicated that they had no training about the CEFR before, either. One prospective teacher discussed the CEFR practices from a more professional perspective and stated:

- "It is a framework which was created by the Council of Europe to draw a general outlook for teaching and assessing process in order to standardise the language teaching studies across Europe. It has a scale from A1 to C2, and each level has descriptors for skills and can-do statements."

The data about what they know about the CEFR were coded as below:

Table 21

A Summary of the Findings Concerning What the Respondents Know about the CEFR

Codes	Number of Respondent	Percentage
has levels	3	75%
standardises language studies	2	50%
by the Council of Europe	2	50%
a framework	2	50%
has can-do statements	1	25%

Two of the instructors stated that they had been informed about the CEFR and its features a few years ago, but they did not remember much, and then with the new administration's effort, they had the chance to learn more in the in-service trainings. As seen in the table, qualitative data's latent content analysis brought up 5 codes. The instructors offering writing course know that the CEFR is "used to standard language studies", and it has "levels". The other two, on the other hand, stated that they know it is " a framework that was created by the Council of Europe, it is the classification of language proficiency in order to determine the levels, it has six levels from A1 to C2, each level has can-do statements."

Research Question 4: "What are the instructors' viewpoints on assessing writing in accordance with the CEFR descriptors?"

On this subject, all of the instructors offering writing course to the focus group students stated that the CEFR is a useful tool providing a common ground for educators. The instructors indicated that they were satisfied with the CEFR-based studies, and would like to see and experience more of them. They indicated that they received assistance from other colleagues while assessing writing and they were always open to suggestions. Two prospective instructors also stated that:

- "The CEFR descriptors enables us to give more objective grades as these descriptors guide us. They describe what to expect from the learners in their writings."
- "The criteria supplied by the CEFR is highly detailed and assistant to define the writing skill requirements for the level in the process. Linguistic competence and skills are described elaborately, which makes it easier to assess tasks objectively."

The instructors are of the same opinions on the benefits of the CEFR descriptors while assessing writing. They believe that the descriptors provided by the CEFR serve the educators on being objective while grading writing tasks. The descriptors are believed to be detailed and assistant while grading writing tasks.

As for the codes on this research question, the data are as the following:

Table 22

A Summary of the Findings Concerning the Respondents' Viewpoints on the Benefits of Assessing Writing in Accordance with the CEFR Descriptors.

Codes	Number of Respondent	Percentage
Objectivity	3	75%
Standardisation	2	50%
Descriptors	4	100%

As seen in the table above, the instructors all believe that the CEFR helps improve writing assessment studies. They indicated the importance of objectivity, standardisation and the descriptors that the CEFR provide. While being interviewed, they repeatedly stated those. Two of the instructors even mentioned their past experiences; the first years of their teaching and the problems they faced while assessing writing. They indicated that they had no problem on that issue any more after the SFL adopted CEFR-based policy.

Research Question 5: "What are the instructors' viewpoints on the contribution of the CEFR to teacher development?"

As in the other questions, the instructors offering writing course have similar ideas on this question as well. They stated that the CEFR helps improve their teaching and assessment ability. While interviewing, one of them pointed out that teaching within the frame of a standardised system contributes him and his students equally. He considered the CEFR as a systematized helping tool for the teachers. The others were of the opinions below:

- "It helps teachers change the idea of teaching; language does not consist of grammar only. The main idea behind language learning is to communicate."
- "The CEFR provides an organized learning environment to the learners and with the guide of the criteria, the teacher can give more reliable grades to the students' writing tasks. Also, following the CEFR criteria provides standardisation among all the teachers."

- "It assists us professionally in every step of teaching process. Since there have never been such elaborate and well-guiding frames like the CEFR, as we use it, we become more conscious and professional about how language teaching was, how it has been and how it should be. Thus, we adopt and modify ourselves to changing current policies."

Table 23

A Summary of the Findings Concerning the Respondents' Viewpoints on the Contribution of the CEFR to Teacher Development

Codes	Number of Respondent	Percentage
Standardisation	4	100%
Criteria	1	25%
Improving teaching	2	50%
Being conscious of assessment	2	50%
Changing the traditional idea of teaching	2	50%

Given the answers above, it is seen that the CEFR is perceived as an assisting and useful tool for teacher development. It is believed to help get more professional in teaching (f=2), standardise assessing process (f=4), be more conscious (f=2), change the traditional idea behind language learning (f=2), which is to communicate, provide necessary criteria and documents.

To compare the responses by the instructors, the codes were grouped into categories as in the table below. Three categories came out of the codes worked on: assessment, benefits of the CEFR and teacher development. The percentage on the "benefits of the CEFR" category is higher than the other two.

Table 24

Categories Concerning the Respondents' Viewpoints on the CEFR Descriptors.

Categories	Codes	Percentage
	Criteria	50%
Assessment	Can-do Statements	25%
	Descriptors	100%
Benefits of the CEFR	Standardisation	100%
	Objectivity	100%
	Gaining experience	50%
	Improving teaching	50%
Teacher Development	Being conscious of assessment	50%
	Changing the traditional idea of teaching	50%

To have a bigger picture on the issue, the categories were grouped into themes as in the table below. Out of the codes and the categories, we have two themes: assessment and professional development. The instructors' responses on the CEFR indicate that two important areas that the CEFR helps us with are these two themes.

Table 25

A Summary of the Findings Concerning the Respondents' Viewpoints on the CEFR

Themes	Categories	Codes
		Criteria
Assessment	Assessment	Can-do Statements
		Descriptors
Professional Development	Benefits of the CEFR	Standardisation
		Objectivity
		Gaining experience
		Improving teaching
	Teacher Development	Being conscious of assessment
		Changing the traditional idea of teaching

Chapter 5
Discussion

With the aim of protecting cultural features and languages of the member countries of the European Union, language policies were bodied by the Council of Europe, and one of the most crucial projects in this area is, of course, the Common European Framework of Reference (Alderson, 2007). This framework aims to guide educators about language learning materials and exams, and to determine language learning goals in the educational institutions and schools of the member countries. The idea behind this project is to come up with a means of language teaching in the member countries so as to compare learners' objectives and achievement standards.

As it is occasionally stated, the CEFR, contrary to what is believed, does not serve educators like a new approach or methodology to language learning and teaching (Council of Europe, 2001). It should be regarded as a tool that can be used to assist teachers, exam developers, course designers and administrators via its practices to be carried out in educational institutions. This requires standardisation and unity among various organizations and education centers. To provide this, the CEFR developed different proficiency levels and their features which indicate language abilities of learners and what they are able to do in the target language. These language proficiency levels are: A1 Breakthrough, A2 Threshold, B1 Waystage, B2 Vantage, C1 Effective Operational Proficiency, C2 Mastery (Council of Europe, 2001).

As Turkey is one of the candidate countries to the EU, the CEFR-based practices have become more important than they were before. Therefore, it is necessary to carry out studies on the CEFR practices so as to render the integration of Turkey to European standards of language teaching as ultimately as possible (Kalkan, 2017). This study was carried out to serve this purpose.

Productive activities and strategies, which are what is required from an autonomous learner, are in the center of a well-organized learning environment in the CEFR practices (Council of Europe, 2001). These activities include both writing and speaking activities. Writing activities are specified in the CEFR handbook, and these activities have been studied in this study as well in order to

have a connection with what is required, how they are required to be performed by the learners and what their roles are in this study.

Another important point is the connection between writing and assessment processes in the CEFR practices. In the third part of the literature review, assessment types are dealt with in an attempt to describe the importance of assessment in CEFR studies.

"Learning to learn", which is stated in the CEFR handbook as "savoir-apprendre", should be an integral element of language learning and to make this happen, learners need to be trained accordingly. In the CEFR, self-assessment grid is of great help giving learners the chance to get more autonomous and assess their own progress. At A2 Level writing classes, the grid goes "I can write short, simple notes and messages relating to matters in areas of immediate need. I can write a very simple personal letter, for example thanking someone for something" (Council of Europe, 2001). With this, learners can set goals and priorities as well as being aware of language learning process. As it is stated in the CEFR (2002), self-assessment and its main potential is in its awareness raising; helping learners to notice their weaknesses. With this thought in mind, the focus group was introduced with a self-assessment questionnaire first and they assessed themselves and their language proficiency, at the end of which it was concluded that they were A2 Level students.

Rubrics help get better and standardised results while assessing writing in accordance with the CEFR. Herein, in this study the rubric that has been used by the instructors in the focus group was compared with the one based on the CEFR descriptors. As seen in the "findings" part, there are a few similarities and differences between these two. The school administrators and course designers tried to adapt their rubric from the CEFR, and they managed to do so. The rubric seems embraced and found well-organized by the instructors. The only problem observed concerning this issue is that there is no separate rubric for different levels at the SFL. The instructors have got used to using just one rubric while grading writing papers at A1, A2 and B1 levels. Having separate rubrics for all levels would help the instructors assess students' papers more properly by taking their differences, levels and abilities into consideration. The instructors grade the

writing tasks in accordance with the rubric they have, and the rubric is not specific to A2 Level, so the instructors grade all the tasks of all levels with the same rubric.

In the second phase of the study, the gradings by the instructors and the standardisation team were examined via SPSS Program to explore if they conform with each other or not. These quantitative findings support the idea that the standardisation among the instructors on grading was achieved. The findings support the current literature. Some of the studies that the findings in this study are consistent with are as follows:

- In the study carried out by Cooper (1984), seventeen experienced teachers were provided a half-day training by the University of Auckland on grading writing tasks of students. Sample scripts, just like the SFL of Karabuk University did, were made use of as a standardisation study after which a cross-checking process was initiated. Consensus coefficient was found as .75.

- In the studies carried out by Supovitz, MacGowan, & Slattery (1997); LeMahieu, Gitomer, & Eresh (1995); Green, Johnson, O'Donovan, & Sutton (2003) in order to look into standardisation in writing assessment, they all attained similar results emphasizing the importance of in-service training.

- The study that was carried out by Chen and Baker (2016) to investigate criterial discourse features by analysing rated learner essays across CEFR B1, B2 and C1 levels. Experienced raters graded the expository and argumentative essays of the students. The stages put into practice in this study are just like the ones in the assessment process of the SFL of Karabuk University: first a familiarization training on the CEFR was performed, then the appropriate samples from essays were selected. 1009 essays were graded by two raters. The inter-rater reliability between two raters was found 0.844.

- Another study carried out by Palviainen, Kalaja and Mantyla (2010) At the University of Jyvaskyla in Sweden. The university students were required to write one narrative and one argumentative text; 41

in Swedish and 62 in English. The texts were assessed by trained raters by using a standard six-point scale based on the CEFR, and to ensure reliability, at least two out of three raters of the texts in Swedish, and three out of four raters of the texts in English had to reach an agreement on the grades. After rating, the relationship between fluency measures and L2 writing proficiency is found not to be inevitably a linear one

Van den Branden (2009) stated that the CEFR is a top-down innovation in the field of education, and it should be supported with teachers' endorsement to be successful in the long term. The qualitative findings obtained through the interview in an attempt to see instructors' perceptions on the CEFR practices revealed that the instructors support the standardisation practices through the CEFR in their institution and it takes all of them to manage this. As Christ (1996) remarked, In order for an innovative initiative to be implemented, education authorities and educators must act in unity. The CEFR practices were found important by the instructors as they provide standardisation and objectivity, and give educators a chance to improve themselves, be conscious of assessment process and have a different teaching idea from traditional ones. The instructors offering writing course know what the CEFR and descriptors mean, what they are intended to do and how they help schools and educators in general. They also believe that the SFL administration has been following the CEFR practices in their studies. The instructors stated that they always make use of the CEFR descriptors, especially while assessing the writing tasks of the students. They do use the CEFR descriptors while assessing writing tasks, and should continue being informed more about the CEFR and its practices as the studies and training in the education institutions should go beyond sharing technical information on it. Other studies such as Papageorgiou (2010), Gad, Ardeshir, Hanan & Chad (2013), Normand-Marconnet (2013), Broek and Ende (2013), Hismanoglu (2013), Ilin (2014), Martyniuk & Noijons (2007), Moonen (2013) and Ngo (2017) also evidenced the positive attitudes of the teachers towards new practices.

Chapter 6
Conclusion and Suggestions
Summary of the Study

This study set out to explore how effectively and to what extent the CEFR descriptors are used while assessing the writing tasks of the A2 Level students at Karabuk University, School of Foreign Languages; therefore a detailed account of the writing assessment system in the school was aimed to be portrayed. As it sought to provide detailed information on the process, the study examined the gradings of the instructors and the standardisation team's in order to see if they conform with each other. Besides, the rubric that the instructors had been using for a while was compared with the one based on the CEFR. Moreover, an interview with the instructors was carried out to shed some light on educators' perceptions of the CEFR practices.

The study adopted a correlation research design to determine if there is a significant difference between the gradings of the instructors and the standardisation team, after which the variables were determined and in the event of a relationship, predictions to the population were made (Simon & Goes, 2011). The scores were compared via SPSS Program. While selecting the group, a nonrandom sampling type, convenience sampling was used (Mackey & Gass, 2005). Furthermore, the rubrics were compared via a comparative and thematic analysis. The data were identified, analyzed, and the patterns were reported within the data (Braun & Clarke, 2006). The interview was used so as to extend more on the quantitative findings. The analysis of the qualitative data was performed through Dörnyei (2007)'s latent content analysis.

Main findings according to research questions are presenten below.

1. The first research question aimed to explore if the writing tasks of the A2 level prep-class students are assessed in accordance with the CEFR descriptors, and as the sub-question, if the rubrics the instructors use are based on the CEFR descriptors. The results revealed that the writing tasks of the A2 level prep-class students are intended to be assessed in accordance with the CEFR descriptors, and the rubric the instructors have been using was tried to be designed according to the CEFR

descriptors, but there are a few differences, and the rubric they have been using is a general one, not specific for A2 Level. The CEFR-based one, on the other hand, was prepared just for A2 Level, which makes it easier to use while grading the writing tasks of the A2 Level students.

2. The aim of the second research question was to explore the extent the instructors' gradings conform with the CEFR standardisation team's gradings. Quantitative data showed that the difference between the grades by the instructors and the ones by the standardisation team is not significant. In the descriptive analysis tables, mean score for instructors for the first task ($\bar{x} =16,82$) is slightly higher than that of standardisation team ($\bar{x} =16,20$). For the second task, mean score for instructors ($\bar{x} =15,92$) is again slightly higher than that of standardisation team ($\bar{x} =15,29$), indicating that there is not a meaningful difference between the mean scores of these two groups.

3. The third research question attempted to investigate what the instructors teaching writing know about the CEFR. Via content analysis, the results revealed that the instructors know the basics about the CEFR; it has levels, it was created by the Council of Europe, it is a framework, it has can-do statements and it standardises language studies.

4. The fourth research question aimed to find out the instructors' viewpoints on assessing writing in accordance with the CEFR descriptors. The codes that were gained via content analysis indicated "objectivity, standardisation and descriptors". The instructors offering writing course all believe that the CEFR helps improve writing assessment studies.

5. The final question sought to provide information about the instructors' viewpoints on the contribution of the CEFR to teacher development, and the results indicated that the CEFR is perceived as an assisting and useful tool for teacher development. The instructors believe that the CEFR helps get more professional in teaching (f=2), standardise assessing process (f=4), be more conscious (f=2), change the traditional idea behind language learning (f=2), which is to communicate, provide necessary criteria and documents.

Conclusion

In today's world, everything is in a state of change including education. Changing needs require integration, standardisation, collaboration and radical transformations from traditional ways of teaching and learning. The CEFR responds to such pedagogical shift as a leading asset in higher education.

This study merely aimed to identify how 93 preparatory students' writing tasks are assessed according to the CEFR descriptors, and what the instructors offering writing course to the focus group know and believe about the CEFR in general. The results of the study confirm previously reported findings and indicated that CEFR practices are there in favour of educators, learners and institutions. The results showed that instructors find these practices useful and effective. The CEFR descriptors were depicted to be guiding, changing the traditional way of teaching and standardising foreign language teaching and learning studies, which were supported with the results gained through the descriptive analysis of gradings by the instructors and standardisation team. The results are believed to provide valuable information for further studies. Yet, more research in this issue is needed as a limited number of participants took part in this research. It can be concluded that the CEFR deserves to be given a chance in every institution as it has the potential to make teaching and learning more effective and standardised.

Implications

The present study offers some pedagogical implications that can be useful for teacher educators, practitioners, administrators and researchers. To begin with, this study may inform them about the CEFR descriptors and how effective they are in writing assessment studies. What makes this study one of the preliminary works for future studies is that it researched what instructors think and know about the CEFR, and compared the gradings by them and the standardisation team like a cross-check. The study, therefore, offers a detailed analysis of the CEFR descriptors in the Turkish context.

Standardisation and unity are two key points that should be paid attention while carrying out CEFR-based practices. The idea behind these practices is to standardise and systematize educational works (Council of Europe, 2002). Given

the results of the study, it seems to be necessary that the instructors should be introduced with separate CEFR-based rubrics designed for each level to provide standardisation. Having just one rubric for all levels might affect the scores the students got from writing task exams; students will be kept being assessed via a rubric designed above or below their level, which brings injustice in scores even if the instructors and the standardisation team grade their writing tasks similarly. All students and teachers should be given this opportunity by their institutions.

Furthermore, as in Leithwood's (1990) Teacher Development Model, at all levels teachers should be a part of the educational decisions. This gives them the highest level in their professional development. The modern teacher profile proposes the involvement of teachers in educational decisions, and the CEFR provides this chance. As the results of the study show, teachers should be informed about the descriptors and approve of the rubrics they are to use in the new academic year.

Another major pedagogical implication of this study is about in-service training. It is an accepted fact that the CEFR practices require systematicity and cooperation among school administrators, teachers, course designers and coordinators (Ahuoglu, 2007). In-service training is considered a good way to manage this since it is regarded as an accelerator to update teachers' knowledge and skills for a better job performance (Omar, 2014). In this study, the instructors who had been offering writing course to the focus group were interviewed to see what they know, believe and think about the CEFR and its practices. The instructors have worked at the SFL for more than 4 years and continuing their master degree studies at different universities, so they have always been a part of the system and seen the progression the school in question has been going through for years. As Berliner stated: "Experience accumulated over years is the most crucial necessary condition for expertise. Nonetheless, mere expertise is certainly not sufficient for it" (Berliner, 1992, p. 161). Therefore, institutions are required to strive hard to train their teachers.

At the very beginning of the first term of the new academic year, every institution willing to carry out the CEFR practices can start by training the teachers as Karabuk University, SFL department did. From the results, the importance of in-

service training is clearly emphasized. For professional development and quality of teaching and learning, in-service training is a crucial tool. Moreover, teachers equip themselves with new skills and knowledge via this type of training. The role of the administrators, teachers' attitude, strategies and needs analysis are influential in terms of the effectiveness of in-service training (Omar, 2014).

Compared with the other studies presented in the literature review part, the results obtained via SPSS program concerning the conformity between the grades by the instructors offering writing course and the standardisation team also state the importance of in-service training. The instructors offering writing course to the prep-class students were trained on the CEFR, descriptors and rubric to be used. The important parts of the process were emphasized and, as the instructors stated, standardisation meetings in which a lot of writing tasks were analyzed in detail and graded by all instructors help them find a common ground. As Ngo (2017) suggested, after in-service training "a robust peer support network" can be created for instructors to discuss and develop CEFR practices in school.

Suggestions for Further Studies

In the light of findings and based on the limitations of the study, some suggestions that may be of great help in the further investigation of the CEFR are presented below:

- The study can be carried out including more students and instructors within a longer period of time in order to get more generalizable results.
- Writing task exam papers from all students all the year around could be analyzed, not just one term.
- A questionnaire aiming to specify students' viewpoints on the assessment studies could be applied, which could also provide a comparison between the views of the instructors and the students.
- Different schools from different regions could be chosen to verify the results and to find out the effects of the CEFR practices on them. By this way, it could be possible to compare the studies and find out if the CEFR is perceived differently in other regions.
- Further research can concentrate on the assessment of other skills.

- Longitudinal investigations into student experiences with the CEFR are strongly recommended.

- Further work with an experimental group (with the CEFR use) and a control group can be included.

- It is recommended that further research could be conducted in different departments at universities.

- More studies on teacher perceptions of the CEFR could be conducted.

- More studies should be conducted to assess the effectiveness of the CEFR practices in the Turkish context.

- Finally, a future study investigating the CEFR descriptors at broader levels at universities would be worthwhile.

References

Affairs, B. E. (2012). *European Language Portfolio.* Council of Europe.

Ahuoglu, S. (2007). Avrupa ortak referanslar çerçevesinde sertifikalar. *Türkiye'de yabancı dil eğitimi ulusal kongresi* (pp. 17-23). Ankara: Gazi Üniversitesi.

Alanen, R., Huhta, A. & Tarnanen, M. (2010). Designing and assessing L2 writing tasks across CEFR proficiency levels. *Communicative Adequacy and Linguistic Complexity in L2 Writing*, 21 (57).

Alderson, J. C. (2005). *Diagnosing foreign language proficiency.* London: Continuum.

Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal, 91*, 658-662.

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

Applebee, A. (1974). *Tradition and reform in the teaching of English: A history.* Urbana, IL: NCTE.

Ariana, S. M. (2010). *Some thoughts on writing skill.* Retrieved from file:///C:/Users/HAMD%C4%B0YE/Downloads/SOME_THOUGHTS_ON_W RITING_SKILLS.pdf

Arslan, A. & Coskun, A. (2012). The Common European Framework of Reference: Turkey. *AIBU Institute of Social Sciences* , 1-19.

Arter, J., Spandel, V., & Culham, R. (1995). *Portfolios for assessment and instruction.* ERIC Digest.

Bachman, L., & Palmer, A. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing 13 (83)*, 257-79.

Bailey, K. (1998). *Learning about language assessment: Dilemmas, decisions, and directions.* Heinle and Heinle: US.

Balliro, L. (1993). What kind of alternative? Examining alternative assessment. *Teachers of English to Speakers of Other Languages (TESOL) Quartely*, 558-561.

Banerjee, J., Franceschina, F., & Smith, A. M. (2004). Documenting features of written language production typical at different IELTS band score levels. *IELTS Research Reports, 7*, 249-309.

Bariskan, V. (2006). *A suggested writing syllabus for students at proficiency level A2-waystage defined in CEFR for languages.* Ankara: Hacettepe University.

Bayraktaroglu, S. (2014). Türkiye'de yabancı dil eğitiminde eğilim ne olmalı? *1. yabancı dil eğitimi çalıştayı bildirileri.* Ankara: Hacettepe Üniversitesi Yayınları.

Bazerman, C. (2008). *Handbook of research on writing: History, society, school, individual, text.* New York: Taylor & Francis Group, LLC.

Berliner, D. (1992). *The nature of expertise in teaching.* San Francisco: Jossey-Bass.

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003). *Assessment for learning: Putting it into practice.* Maidenhead: Open University Press.

Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2004). *Working inside the black box: Assessment for learning in the classroom.* Slough: NFER Nelson.

Boud, D. (1995). *Enhancing learning through self-assessment.* London: Kogan Page.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 77-101.

Brindle, M. & Harris, K. (2016). Third and fourth grade teacher's classroom practices in writing: A national survey. *Reading and Writing 29*, 929-954.

Brindley, G. (2009). *Task-centered language assessment in language learning. The Promise and the challenge. In Norris, J., Bygate, M. & Van den Branden, K. (eds.), Task-based language teaching.* Amsterdam & Philadelphia: Benjamins.

Broek, S. & Ende, I.V.D. (2013). *The implementation of the Common European Framework for languages in European education systems.* Retrieved from http://www.europarl.europa.eu/studies.

Brookhart, S. & Chen, F. (2014). The quality and effectiveness of descriptive rubrics. *Educational Review*, 1-26.

Broughton, G., Brumfit, C., Flavell, R., Hill, P. & Pincas, A. (2003). Writing. In *teaching English as a foreign language* (pp. 116-132). London: Routledge.

Buhagiar, M. (2007). Classroom assessment within the alternative assessment paradigm: Revisiting the territory. *The Curriculum Journal 18 (1)*, 39-56.

Butler, Y. & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language testing 27 (1)*, 5-31.

Campbell, A., McNamara, O., & Gilror, P. (2004). *Practitioner research and professionel development in education.* London: Paul Chapman.

Campbell, R. & R, Wales. (1970). *The Study of language acquisitios. In Lyons, J. (eds). New Horizons in Linguistics.* Harmondsworth: Penguin Books.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 3-8.

Carlsen, C. (2010). Discourse connectives across CEFR-levels: A corpus based study. *Eurosla Monographs Series 1*, 191-207.

Celce-Murcia, M. (2001). *Teaching English as a second or foreign language. 3rd Ed.* Heinle & Heinle Publishers.

Celce-Murcia, M., Dörnyei, Z. & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 7.

Chen, Y. &. (2016). Investigating criterial discourse features across second language development:Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics 37/6*, 849-880.

Christ, H. (1996). Didaktische konzepte im umfeld der sprachenpolitik am ende des 20. jahrhunderts. Funk, H. & Neuner, G. (Hg.). *Verstehen und Verständigung in Europa*, 182-192.

Christensen, L. B. (2015). *Research methods, design, and analysis.* Essex: Pearson Education Limited.

Coaley, K. (2010). *An introduction to psychological assessment and psychometrics.* London: Sage Publications.

Coffman, W. (1971). On the reliability of ratings of essay examinations in English. *Research in the Teaching of English, 5*, 25-34.

Connor, U. (2002). *New direction in contrastive rhetoric.* TESOL Quarterly.

Cooper, P. (1984). *The assessment of writing ability: A review of research (ETS Research Report No: 84-12).* Princeton, NJ: Educational Testing Service.

Coste, D. (2007). *Contextualising uses of the Common Eurpean Framework of Reference for languages.* Strasbourg: Council of Europe.

Council of Europe. (1997). *Language learning for European citizenship: Final Report of the Project.* Strasbourg: Council of Europe.

Council of Europe. (2000). *European Language Portfolio (ELP) principles and guidelines.* Language Policy Division.

Council of Europe. (2001). *A Common European Framework of Reference for languages: Learning, teaching, assessment.* Cambridge: CUP.

Council of Europe. (2002). *Common European Framework of Reference for languages: Learning, teaching, assessment.* Strasbourg: Council of Europe.

Council of Europe. (2002). *Common European Framework of Reference for languages: Learning, teaching, assessment: Case studies.* Strasbourg: Council of Europe.

Council of Europe. (2003). *Manual for relating examinations to the Common European Framework of Reference for languages.* Strasbourg: Council of Europe.

Council of Europe. (2008). *Recommendation of the committee of ministers to member states on the use of the Council of Europe's Common European Framework of Reference for languages (CEFR) and the promotion of plurilingualism.* Strasbourg: Council of Europe.

Council of Europe. (2009). *Manual for relating language examinations to the Common European Framework of Reference for languages (CEFR).* Retrieved from Council of Europe: www.coe.int/t/dg4/linguistic/

Council of Europe. (2018). *European Language Portfolio.* Retrieved from Council of Europe: https://www.coe.int/en/web/lang-migrants/european-language-portfolio-elp-l

Creswell, J. W. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: SAGE Publications.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* Boston, MA: Pearson.

Daghan, G. & Akkoyunlu, B. (2014). A qualitative study about performance based assessment methods used in information technologies lesson. *Educational Sciences: Theory and Practice 14 (1)*, 333-338.

Deniz, S. & Sahin, N. (2006). The restructuring process of teacher training system in Turkey: A model of teacher training based on post-graduate education (PGCE). *Journal of Social Sciences 2*, 21-26.

Denscombe, M. (2008). Communities of practice: A research paradigm for the mixed methods approach. *Journal of Mixed Methods Research, 2(3)*, 270-283.

Diederich, P. B. (1961). Factors in judgements of writing ability. (pp. 61-15). Princeton, NJ: Educational Testing Service.

Dietel, R., Herman, J., & Knuth, R. (1991). *What does research say about assessment?* Oak Brook: NCREL.

Dochy, F., Segers, M. & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Educational Evaluation*, 50-331.

Dogan, M. (2011). Student teachers' views about assessment and evaluation methods in mathematics. *Education Research Review 6 (5)*, 417-431.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies.* Madrid: Oxford University Press.

DuBois, F. (1970). *A History of psychological testing.* Boston: Allyn & Bacon.

Durán, P. & Pierce, J. (2010). *Self-assessment based on language learning outcomes: A study with first year engineering students.* Revista Alicantina de Estudios Ingleses 23.

ECTS Users' Quide. (2009). *Luxembourg: office for official publications of the European communities.* Retrieved from http://ec.europa.eu/education/lifelong-learning-policy/doc/ects/quide_en.pdf

Edgeworth, F. Y. (1888). The statistics of examinations. *Journal of the Royal Statistical Society, 51*, 599–635.

Ekbatani, G. (2000). Moving toward learner-directed assessment. In *Learner-directed assessment in ESL* (pp. 1-11). Mahwah, NJ: Erlbaum.

Elliott, S. (1995). *Creating meaningful performance assessments.* ERIC Digest E531.

European Commission. (2016). *Turkey report of European Commission. Bruxelles.* Retrieved from www.ab.gov.tr: http://www.ab.gov.tr/files/5%20Ekim/son__2016_ilerleme_raporu_tr.pdf

Forsberg, F. & Bartning, I. (2010). Can linguistic features discriminate between the communicative CEFR-levels? A pilot study of written L2 French. *Eurosla Monographs Series 1*, 133-158.

Fourie, I. & Niekerk, D. V. (2001). Follow-up on the portfolio assessment. A module in research information skills: An analysis of its value. *Education for information 19 82)*, 107-126.

Fraenkel, J. R. (2012). *How to design and evaluate research in education.* New York: McGraw-Hill.

Fulcher, G. (2008). Testing times ahead? *Liaison Magazine (1).*

Gad, S. L. (2013). Standard setting to an international reference framework: Implications for theory and practice. *International Journal of Testing, 13(1)*, 32.

Garcia, G. & Pearson, P. (1994). *Assessment and diversity. In Darling Hammond, L.(Ed.), Review of research in education.* Washington, DC.: American Education Research Association.

Gearhart, M., Herman, J., Novak, J., Wolf, S. & Abedi, J. (1994). *Toward the instructional utility of large-scale writing assessment: Validation of a new narrative rubric.* Los Angeles, CA: CRESST.

Gill, D. & Lucas, D. (2013). Using alternative assessment in business and foreign language classes. *Journal of International Education Research 9 (4)*, 359-370.

Gipps, C. (2012). *Beyond testing: Towards a theory of educational assessment, Classic Edition.* Abingdon: Routledge.

Glasswell, K., Brown, G. & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, 105-121.

Glover, P. & Mirici, I. H. & Aksu, M. B. (2005). Preparing for the European Language Portfolio: *The Turkish Online Journal of Distance Education*, 84-98.

Green, S., Johnson, M., O'Donovan, N. & Sutton, P. (2003). Changes in key stage two writing from 1995 to 2002. *The United Kingdom Reading Association Conference.* Cambridge, UK.

Gronlund, N. & Linn, R. (1990). *Measurement and evaluation in teaching (6th Ed.).* New York: Macmillan.

Hamp-Lyons, L. (1991). *Assessing second language writing in academic contexts.* Norwood, N.J: Ablex Pub.Corp.

Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing 8*, 5-16.

Harsch, C. & Rupp, A. A. (2011). Designing and scaling level-specific writing tasks in alignment with the CEFR: A test-centered approach. *Language Assessment Quarterly, 8 (1),* 1-33.

Harsch, C. (2007). *Der gemeinsame Europäische referenzrahmen für sprachen: leistung und grenzen (The CEFR for languages: Strengths and limitations).* Saarbrücken, Germany: VDM.

Hartog, S. P. (1910). *Examinations. Encyclopaedia Britannica, 1910 Ed.* Cambridge: Encyclopaedia.

Hayes, J.R. & Berninger, V. (2014). *Cognitive processes in writing: A framework. In Arfe, B., Dockrell, J. & Berninger, V. (Eds.), Writing development and instruction in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction (pp. 3-15).* Oxford, UK: Oxford University Press.

Herman, J. L. (1992). *A practical guide to alternative assessment.* Association for Supervision and Curriculum Development. Washington, DC.

Heyworth, F. (2005). The Common European Framework. *Metro Xpress Bulletin*, 8-11-19.

Hismanoglu, M. (2013). Does English language teacher education curriculum promote CEFR awareness of prospective EFL teachers? *Procedia - Social and Behavioral Sciences, 93(0)*, 938-945. http://dx.doi.org/10.1016/j.sbspro.2013.09.307.

Hoghes, A. (2003). *Testing for language teachers (2nd Ed.).* Cambridge: Cambridge University Press.

Holec, H. (1990). What is learning to learn? *Melanges Pedagogiques*, 75-87.

Hook, J. N. (1979). *A long way together.* Urbana, IL: NCTE.

Houge, A. (2007). *First steps in academic writing. 2nd Ed.* New York : Pearson Longman.

Hughes, A. (1989). *Testing for language teachers.* Cambridge: Cambridge University Press.

Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal, 91*, 663-667.

Humphry, S. & Heldsinger, S. (2014). Common structural design features of rubrics may represent a threat to validity. *Educational Researcher 43 (5)*, 253-263.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research, 60 (2)*, 237–263.

Hymes, D. (1972). *On communicative competence. In pride, J.B. & Holmes, J. (Eds.), Sociolinguistics, 269-293.* Harmondsworth: Penguin Education, Penguin Books Ltd.

Ilin, G. (2014). Student-teacher judgements on Common European Framework: efficacy, feasibility and reality. *Journal of Language and Literature Education 2 (9)*, 8.

Isaacs, T., Zara, C., Herbert, G., Coombs, S.J. & Smith, C. (2013). *Key concepts in educational assessment.* Sage Publications. Available at https://ebookcentral.proquest.com/lib/ucl/reader.action?docID=1138514&ppg=13.

Janesick, V. (2001). *The assessment debate: A reference handbook.* Santa Barbara: CA.

Janisch, C., Liu, X. & Akrofi, A. (2007). Implementing alternative assessment: opportunities and obstacles. *The Educational Forum 71.*, 221-230.

Janssen-van Dieten, A. (2003). *The European Framework of Reference and L2 learners with a low level of education.*

Johnson, R. B. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher 33 (7)*, 14-26.

Kalkan, E. (2017). *Avrupa Dilleri Ortak Çerçeve Programı kapsamında Türkiye'de yabancı dil öğretmeni yetiştirme politikaları*. Retrieved from www.academia.edu:https://www.academia.edu/34097608/AVRUPA_D%C4%B0LLER%C4%B0_ORTAK_%C3%87ER%C3%87EVE_PROGRAMI_KAPSAMINDA_T%C3%9CRK%C4%B0YE_DE_YABANCI_D%C4%B0L_%C3%96%C4%9ERETMEN%C4%B0_YET%C4%B0%C5%9ET%C4%B0RME_POL%C4%B0T%C4%B0KALARI.pdf?auto=download

Kalra, R. (2017). Using portfolio as an alternative assessment tool to enhance Thai EFL students' writing skill. *Arab Worl English Journal 8 (4)*, 292-302.

Kaplan, R. (1996). *Cultural thought patterns in intercultural education. Language Learning XVI (1&2).*

Kelchtermans, G. & Vandenburghe, R. (1994). Teachers' professional development: A biographical perspective. *Journal of Curriculum Studies 26(1)*, 45-62.

Koc, B.B., Alparda, C., Kulac, D., Bicer, D., Baki, E., Saka, E. & Kose, E. (2008). *Milestones of writing.* Zonguldak: Karaelmas University.

Koreltz, D. (2008). *Measuring up: What educational testing really tells us.* Cambridge, MA: Harvard University Press.

Krashen, S. (1982). *Principles and practices in second language acquisition.* Oxford: Pergamon.

Kubota, R. & Lehner, A. (2004). Toward Critical Contrastive Rhetoric. *Journal of Second Language Writing*, 7-27.

Kulieke, M., Collins, C., Fennimore, T., Fine, C., Herman, J., Jones, B. & Tinzmann, M. (1990). *Why should assessment be based on a vision of learning?* NCREL, Oak Brook.

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing.* Thousand Oaks, CA: Sage.

Law, B. & Eckes, M. (1995). *Assessment and ESL.* Manitoba, Canada.: Peguis publishers.

Leithwood, K. (1990). *The principal's role in teacher development.* Virginia : ASCD.

LeMahieu, P., Gitomer, D. & Eresh, J. (1995). Large-scale portfolio assessment: Difficult but not impossible. *Journal of Educational Measurement: Issues and Practice, 14*, 11-28.

Lenz, P. & Schneider, G. (2004). *A bank of descriptors for self-assessment in European Language Portfolios.* Fribourg: Council of Europe, Language Policy Division.

Linn, R. & Gronlund, N. (2000). *Measurement and evaluation in teaching (8th Ed.).* New York: Macmillan.

Little, D. & Perclova, R. (2001). *European Language Portfolio: a guide for teachers and teacher trainers.* Strasbourg: Council of Europe. Available at https://www.coe.int/en/web/portfolio/elp-related-publications (January 10, 2018).

Little, D. (2005). The Common European Framework and the European Language Portfolio: Involving learners and their judgments in the assessment process. *Language Testing*, 22/3, 320-336.

Little, D. (2006). The CUFRL: Content, purpose, origin, reception, and Impact. *Language Teaching 39*, 167-190.

Little, D. (2007). The Common European Framework of Reference for languages: Perspectives on the making of supranational language education policy. *Modern Language Journal 91(4)*, 645-653.

Little, D. (2008). Knowledge about language and learner autonomy. In Cenoz, J. & Hornberger, N.H. (Eds.). *Encyclopedia of language and education: Knowledge about language 6*, 247-258.

Little, D. (2009). The European Language Portfolio: where pedagogy and assessment meet. *8th International Seminar on the European Language Portfolio* (pp. 1-17). Graz: Council of Europe. Available at https://www.coe.int/en/web/portfolio/reports.

Little, D. (2015). Making good use of the CEFR: Some reflections on the PRO-signs projects.

Lloyd-Jones, R. (1977). *Primary trait scoring. In Cooper, C.R. & Odell, L. (Eds.), Evaluating Writing 33-69.* NY: National Council of Teachers of English.

Lombard, M. S. D. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28(2)*, 587-604 doi:10.1111/j.1468-2958.2002.tb00826.x.

Lunsford, A. (1986). *The past & future of writing assessment. In: Greenberg, K. L., Wiener, H. S. & Donovan, R. A. (Eds.), Writing assessment: Issues and strategies (pp. 1–12).* New York: White Plains, Longman.

Macaro, E. (2003). *Teaching and learning a second language.* London: Continuum.

Mackenzie, N. (2013). Analysing writing: The development of a tool for use in the early years of schooling. *Issues in Educational Research 23 (3)*, 375-393.

Mackey, A. & Gass, S.M. (2005). *Second language research: Methodology and design.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Martyniuk, W. & Noijons, S. (2007). *Executive summary of results of a survey on the use of the CEFR at national level in the Council of Europe member states.* Retrieved from http://www.coe.int/t/dg4/linguistic/Source/Survey_CEFR_2007_EN.doc.

Martyniuk, W. (2006). *Common European Framework of Reference for language competences.* Strasbourg: Council of Europe.

Martyniuk, W. (2006). Common European Framework of Reference for languages: Learning, teaching, assessment (CEFR) - a synopsis. *Annual meeting of the consortium for language teaching and learning. Cornell University*.

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach.* California, CA: Sage Publications.

McMillan, J. (2001). *Classroom assessment: Principles and practice for effective instruction (2nd Ed.).* Boston, MA: Allyn & Bacon.

McNamara, T. (1996). *Measuring second language performance.* London and New York: Longman.

Miles, M. B. (1994). *Qualitative data analysis: An expanded source book (2nd Ed.).* Thousand Oaks, CA: Sage.

Mirici, İ. H. (2017). *Avrupa Dilleri Öğretimi Ortak Çerçeve Programı ve Avrupa Dil Portfolyosu nedir?* Retrieved from www.meb.gov.tr: http://adp.meb.gov.tr/nedir.php

Moonen, M. S. (2013). Implementing CEFR in secondary education: Impact on FL teachers' educational and assessment practice. *International Journal of Applied Linguistics, 23 (2)*, 226-246 doi: 10.1111/ijal.12000.

Moreno, J. (2003). *Brief information: The Common European Framework of Reference for languages and the European Language portfolio.* Retrieved from http://www.ecml.at/documents/pub214E2003Boldizsar.pdf

Myford, C. & Wolfe, E. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: part I. *Journal of Applied Measurement 4 (4)*, 386-422.

Ngo, X. (2017). Diffusion of the CEFR among Vietnamese teachers: A mixed methods investigation. *Asian EFL Journal* , 7-32.

Normand-Marconnet, N. & Lo Bianco, J. (2013). Importing language assessment? The reception of the Common European Framework of Reference in Australian universities. *European Conference on Language Learning 2013.* www.iafor.org/offprints/ecll2013.../ECLL2013_Offprint_0468.pdf.

North, B. (1993). The Development of descriptors on scales of language proficiency: Prospectives, problems and a possible methodology based on a theory of measurement. *NFLC Occasional Papers*.

North, B. (1994). *Scales for language proficiency. A survey of some existing systems.* Strasbourg: Council of Europe.

North, B. (1995). The Development of a Common Framework scale of descriptors of language proficiency based on a theory of measurement. *Pergamon*, 445-465.

North, B. (2014). Putting the Common European Framework of Reference to good use. *EAQUALS* (pp. 228-244). Switzerland: Eurocentres Foundation.

Nunan, D. (1988). *The Learner-centred curriculum.* Cambridge: Cambridge University Press.

Oliver, E. (2015). Alternative assessment to enhance theological education. *HTS Teologiese Studies / Theological Studies 71 (3)*, 1-10.

Olshtain, E. (2001). Functional tasks for mastering the mechanics of writing and going just beyond. In Celce-Murcia, M. *Teaching English as a Second or foreign language* (pp. 207-213). Heinle & Heinle Publishing.

Omar, C. (2014). The need for in-Service training for teachers and its effectiveness in school. *International Journal for Innovation Education and Research*, pp. 1-7.

Oproescu, A. (2018). The role of alternative assessment methods on school performance. *ECAI 2018 International Conference 10th Edition. Computers and Artificial Intelligence*, 1-5.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing 6 (1)*, 1-13.

Oshima, A. & Hogue, A. (2007). *Longman academic writing series 4th Ed. Paragraphs to essays.* New York: Pearson.

ÖSYM. (2018). *Student selection and placement guide.* Retrieved from Higher Education Council: www.osym.gov.tr

Pachler, N. & Redondo, A. (2006). *A practical guide to teaching modern foreign languages in the secondary school.* London : Routledge (Taylor and Francis Group).

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30*, 590-601.

Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research & Evaluation 13 (4)*, 1-11.

Palviainen, A., Kalaja, P. & Mantyla, K. (2012). Development of L2 writing: fluency and proficiency. *AFinLA-e Soveltavan kielitieteen tutkimuksis 2012 / n:o 4.*, 47-59.

Papageorgiou, S. (2010). Investigating the decision-making process of standard setting participants. *Language Testing, 27 (2)*, 261-282 doi: 10.1177/0265532209349472.

Paris, S. & Paris, A. (2001). Classroom applications of research on self-regulated learning. *Educational Psychology 36 (2)*, 89-101.

Paterson, D. (1925). *Preparation and use of new-type examinations: A manual for teachers.* Orlando, FL: Harcourt Brace Jovanovich.

Patton, M. Q. (2002). *Qualitative research and evaluation methods (3rd Ed.).* Thousand Oaks, CA: Sage Publications.

Penny, J., Johnson, R. & Gordon, B. (2000). The Effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing, 7*, 143-164.

Perakyla, A. & Ruusuvuori, J. (2011). Analyzing talk and text. In Denzin, N.K. & Lincoln, Y.S. (Eds.). *The SAGE handbook of qualitative research*, pp. 529-543.

Glover, P., Mirici, İ.H. & Aksu, M. B. (2005). Preparing for the European Language Portfolio: Internet connections. *The Turkish Online Journal of Distance Education*, 84-98.

Piccardo, E., Berchoud, M., Cignatta, T., Mentz, O. & Pamula, M. (2009). *Pathways through assessing, learning and teaching in the CEFR.* Strasbourg: Council of Europe Publishing.

Poon, A. (2004). *Action research: A study on using an integrative-narrative method to teach L2 writing in a Hong Kong primary school.* Hong Kong : Hong Kong Baptist University.

Rashdall, H. (1895). *The universities of Europe in the middle ages.* London, Deighton: Bell & Co.

Rasheed, M. I., Aslam, H. D. & Sarwar, S. (2010). Motivational issues for teachers in higher education: A critical case of IUB. *Journal of Management Research 2 (2)*, 1-23.

Resnick, L., Resnick, D. & De Stefano, L. (1993). Cross-scorer and cross-method comparability and distribution of judgments of student math, reading and writing performance: Results from the New Standards Project. *Big sky scoring conference.* Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Rezaei, A. & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing 15 (1)*, 18-39.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experimental factors. *Language Testing 15 (1)*, 1-19.

Savignon, S. (1976). *Communicative competence: Theory and classroom practice.* University of Illinois: Urbana-Champaign.

Scharer, R. (2000). *European Language Portfolio: Final report on the pilot project.* Strasbourg: Council of Europe. https://www.coe.int/en/web/portfolio/reports.

Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences.* New York: Teachers College Press.

Shavelson, R. & Webb, N. (1991). *Generalizability theory: A primer.* Newbury Park, CA: Sage.

Shepard, L. (2000). The role of assessment in a learning culture. *Education Research 29 (7)*, 4-14.

Shermis, M. (2014). State of the art automated essay scoring: Competition, results and future directions from a United States demonstration. *Assessing Writing 20 (*6), 53-76.

Shneider, G. & Lenz, P. (2006). *European Language Portfolio guide for developers.* CH: University of Fribourg.

Silverman, D. (2000). *Doing qualitative research.* London: Sage.

Simon, M. & Goes, J. (2011). *Correlation research.* Seattle: Dissertation Success LLC.

Simonson, M., Smaldino, S., Albright, M. & Zvacek, S. (2000). Teaching and learning at a distance: Foudations of distance education. In: *Assessment for distance education (ch. 11).* Upper Saddle River, NJ: Prentice-Hall.

Spandel, V. (2005). *Creating writers through 6-trait writing: Assessment and instruction (4th Ed.).* Boston: Pearson and Allyn & Bacon.

Spolsky, B. (1996). *The examination-classroom backwash cycle: Some historical cases. In: Nunan, D., Berry, R. & Berry, V. (Eds.), Bringing about change in language education (pp.55-66).* Hong Kong: Department of Curriculum Studies: The University of Hong Kong.

Stefani, L. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Educational Evaluation 19 (1)*, 60-75.

Stemler, S. (2004). A comparison of consensus, consistency and measurement approaches to estimating inter-rater reliability. *Practical Assessment Research & Evaluation, 9 (4)*.

Stiggins, R. (2001). *Student-involved classroom assessment (3rd Ed.).* Columbus, OH: Merrill Prentice Hall.

Strauss, A. & Corbin, J. (1990). *Basics of qualitative research.* Newbury Park, CA: Sage.

Stuhlmann, J., Daniel, C., Dellinger, A., Denny, R. & Powers, T. (1999). A generalisability study of the effects of training on teachers' abilities to rate children's writing using a rubric. *Reading Psychology, 20 (2)*, 107-127.

Supovitz, J., MacGowan, A. & Slattery, J. (1997). Assessing agreement: Interrater reliability of portfolio assessment in Rochester, New York. *Educational Assessment, 4 (3)*, 237-259.

Tan, K. (2004). Does student self-assessment empower or discipline students? *Assessment and Evaluation in Higher Education, 29 (6)*, 62-651.

Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutor. *Assessment and Evaluation in Higher Education 26 (6)*, 14-605.

Teddlie, C. & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences.* Thousand Oaks, CA: Sage Publications.

The Impel Glossary. (2007). *A glossary of European Language Portfolio (ELP) Implementation Terminology Version.*

Thomas, C. S. (1931). *Examining the examinations in English: Report to the college entrance examination board.* Cambridge, MA: Harvard University Press.

Van den Branden, K. (2009). *Diffusion and implementation of innovations. In Long, M. H. & Doughty, C. J. (Eds.), The handbook of language teaching (pp. 659-672).* Hoboken, NJ: Wiley-Blackwell.

Van Ek, J. A. & Trim, J. L. M. (1998). *Waystage.* Cambridge: CUP.

Van Ek, J. A.& Trim, J. L. M. (1991). *Waystage 1990 Council of Europe.* New York: Cambridge University Press.

Weigle, S. C. (2002). *Assessing writing.* New York: Cambridge University Press.

Weir, C. (1990). *Communicative language testing .* NJ: Prentice Hall Regents.

Weir, C. (2005). *Language testing and validation: An evidence-based approach.* London: Palgrave Macmillan.

Wernicke, M. (2014). Action-oriented language teaching "ja genau!". *Forum Deutsch: Forschungsforum, 22 (1)*, 1-21.

West, R., Guven, A. & Parry, J. (2015). *Türkıye'de yükseköğretim kurumlarında Ingilizce eğitimi.* Ankara: British Council.

White, E. (1984). Holisticism. *College Composition and Communication 35 (4)*, 400-409.

Wiggins, G. (1998). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan 70 (9)*, 703-713.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37 (1)*, 3-14.

Wiseman, S. (1956). Symposium on the use of essays on selection at 11+. *British Journal of Educational Psychology, 26 (3)*, 172-179.

Wright, R. (2007). Educational assessment: Tests and measurements in the age of accountability. New York: Sage Publications.

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication, 50 (3)*, 483-503.

Zemach, D.E. & Rumisek, L.A. (2005). *Academic writing: From paragraph to essay.* Thailand: Macmillan.

An Ordinary Day

Topic Sentence: Last Ramadan was an ordinary feast for me.

1. First Day (Major Support)
    A. celebrating each other (Minor Support)
    B. visiting our relatives Minor Support)
    C. meeting friends (Minor Support)
    D. visiting famous places (Minor Support)
2. Second Day (Major Support)
    A. watching TV at home (Minor Support)
    B. relatives coming to visit us in the evening (Minor Support)
3. Last Day (Major Support)
    A. doing my homework and some revisions (Minor Support)

Concluding Sentence: In short, the last Ramadan was as usual as always.

(Taken from Koc, B.B. et al., 2008, p. 78)

An Ordinary Day

Last Ramadan was an ordinary feast for me. First, my family and I got up early and celebrated each other. In the afternoon, we visited our relatives. This was very boring. Then, I met my friends at a café. We talked about our school and social lives. Next, we visited some places famous worldwide such as the Topkapı Palace and the Hagia Sophia Mosque. On the second day, I was at home. I watched television, had some rest and read some books. In the evening, our relatives visited us. They brought some chocolate and candies with them. On the last day, I did my homework and some revision of the latest subjects at school. In short, the last Ramadan was as usual as always.

(Taken from Koc, B.B. et al., 2008, p. 78)

My Dream Partner

Topic Sentence: My dream partner is the most wonderful girl in the world.

I. physical appearance
    A. hair
        1. soft, wavy, dark
        2. clean and bright
    B. oval face
        1. arched eyebrows
        2. charming blue eyes
        3. well-shaped nose
        4. cheeks as red as a nose
        5. full lips
        6. round chin
    C. slim body
II. Character
        1. generous
        2. friendly
        3. successful and popular actress

Concluding Sentence: All in all, her pure beauty and unique character, she is the woman of my dreams

(Taken from Koc, B.B. et al., 2008, p. 90)

My Dream Partner

My dream partner is the most wonderful girl in the world. First of all, her physical appearance is marvelous. She has soft, wavy, dark hair. Her hair is always clean and bright. On her oval face the first outstanding feature is her arched eyebrows. Under her eyebrows, you can immediately recognize her charming blue eyes. They take you to the deep blue sea. There is so little make-up on them, this makes her seem more natural. Her well-shaped nose reflects her strong character. As for her cheeks, they are as red as arose. Her lips are full and there is always a childish expression on her face when she smiles. Her round chin matches well with her beautiful face. Also, her wonderful posture makes her slim body look even more beautiiful. Secondly, beside her beauty, her character is perfect. She is so generous. She always helps poor people. She is quite friendly and she gives positive energy to the people around her. Also, she has a God-given talent. This makes her a successful and popular actress. All in al, her pure beauty and unique character, she is the woman of my dreams

(Taken from Koc, B.B. et al., 2008, p. 90)

APPENDIX - C: "Can Do" Statements

My Linguistic and Intercultural Attainments   Dilsel ve Kültürlerarası Kazanımlarım

My Personal Language Attainments   Kişisel Dil Kazanımlarım

Please specify your personal language attainments below.

Lütfen kişisel dil kazanımlarınızı aşağıda belirtiniz.

Please use this arrow to reflect what you think you can do as given in the example.

Lütfen yapabildiğinizi düşündüğünüz dil becerilerinizi ok üzerinde örnekteki gibi belirtiniz.

RARELY in the box for Level 1,

NADİREN ise Düzey 1

1	2	3	4

SOMETIMES in the box for Level 2,

BAZEN ise Düzey 2,

1	2	3	4

OFTEN in the box for Level 3,

SIKLIKLA ise Düzey

1	2	3	4

ALWAYS in the box for Level 4

HER ZAMAN ise Düzey 4

1	2	3	4

1) … using basic expressions and very simple sentences.

Basit sözcük grupları ve çok basit cümleler kullanarak …
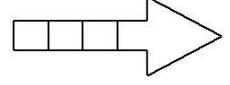
2) I can create notes about where I am, or where and

   when to meet somebody

nerede olduğum, ya da birisiyle buluşma yeri ve zamanı hakkında bir not oluşturabilirim.
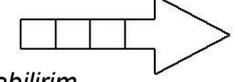
3) I can write a short letter to express my thanks, to apologise,

   to send greetings

teşekkürlerimi aktarmak, özür dilemek, selamlarımı iletmek için kısa bir mektup yazabilirim
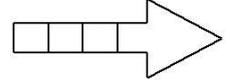
4) I can describe my family, myself, my hobbies, my school or my job in short texts

ailemi, kendimi, hobilerimi, işimi ya da okulumu kısa metinlerde betimleyebilirim.
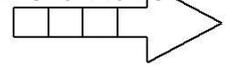
5) I can describe an event or an activity such as a celebration

   or a meeting in simple sentences

kutlama ya da toplantı gibi bir olayı ya da etkinliği basit cümlelerle anlatabilirim
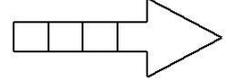
6) I can write short stories using pictures

resimler kullanarak kısa hikayeler yazabilirim
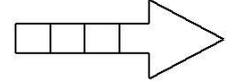
7) I can write short texts with introduction, development and conclusion parts,

using sentence connectors such as "first", "then", "after that" and "because"

"Önce", "sonra", "daha sonra" ve "çünkü" gibi bağlaçlar kullanarak giriş, gelişme ve

sonuç bölümü olan kısa metinler yazabilirim.
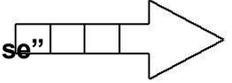
8) I can fill in a form about my educational background, my job,

my fields of interest and my specific skills

eğitim durumum, işim, ilgi alanlarım ve özel becerilerim hakkında bir form doldurabilirim.

(Taken from www.bedaf.org)

APPENDIX - D: A Writing Task Sample

KBU SCHOOL OF FOREIGN LANGUAGES
2017-2018 MODULE 2/ PERIOD 2 WRITING TASK 1

Name-Surname: Aslı Erdem   Class: C-4
Number: 2017010446056

Think of a special occasion (a birthday party, a wedding ceremony, graduation, religious festival etc.) you celebrated with your friends or family members. Write a paragraph to describe it using *at least 80 words*. You can give information about who you were with, when it was, what you did, how you felt, etc. Do not forget to **write a title** for your paragraph.

My Special Day

My special day is graduation. I got up early two years ago. My family and I had breakfast together. I went to school at eight o'clock. I felt happy. I felt sad. I saw my friends in my class. I talked my teachers and my friends. I went to my old school at twelve o'clock. I talked my old friends. I said "See you four years later." We were hug. Then, I went back my house. I had dinner at seven o'clock. I went to the graduation party. We ate cake. We drunk someone? We had dance. I felt sad in the end of the night. We were hug. I left my friends. I went to my house. I slept. Graduation is my special day because it is end of the high school.

Content (5)	Organization (5)	Vocabulary(4)	Grammar (4)	Mechanics (2)	Total (20pts)
4	4	4	3	2	17

Session 1

APPENDIX - E: The Witing Task Assessment Rubric with the CEFR Descriptors

	0	1	2	3	4	5
Content	• Hardly any or no relevance to the task • Hardly any or no valid ideas in completing the task	• Few content points mentioned • Few valid ideas and/or iterative ones in completing the task	• Some content points mentioned • Some valid but insufficient ideas in completing the task	• Several content points mentioned • Reasonable attempt to have sufficient, valid ideas in completing the task	• Most content points mentioned • Sufficient, valid ideas in completing the task	• All content points fully mentioned • Wide range of valid , relevant ideas in completing the task
Organization	• Completely disorganized • No unity, coherence and logical sequencing	• Considerably disorganized • Mostly incoherent; ideas disconnected • Incorrect or rare use of simple linking devices	• Inadequately organized ideas • Considerably incoherent; ideas are not well connected	• Adequately organized valid ideas • Mostly coherent and fluent • Some simple linking devices	• Clearly well-organized valid ideas • Coherent and fluent • Appropriate linking devices	• Effectively organized valid ideas • Very coherent and fluent • A wide range of appropriate linking devices
Vocabulary	• Very poor / insufficient knowledge of basic vocabulary • Irrelevant words	• Narrow, inadequate range of words • Translation-based errors • Mistakes in using some simple, short everyday chunks	• Limited range of vocabulary choice that is relevant to the content • Mistakes in linking groups of words with simple connectors	• Good range of appropriate vocabulary • Some mistakes of word, but no obscure in communication • Few mistakes in linking groups of words with simple connectors	• A wide range of appropriate vocabulary • Hardly any mistakes in the word choice • Hardly any mistakes in linking groups of words with simple connectors	
Grammar	• Serious lack of some simple language structures	• Inadequate range of structures • The language is rarely clear • Frequent mistakes in the use of simple structures	• Limited range of structures • Mistakes in forming simple sentences • Occasional mistakes; mostly interference from mother tongue	• Good range of appropriate structures • Generally accurate language • Few mistakes in forming simple sentences and phrases	• A wide range of appropriate structures • Simple phrases and sentences linked well with simple connectors • Systematically basic mistakes	
Mechanics	• No control over spelling and punctuation, capitalization, and paragraph indentation etc,	• Occasional mistakes in spelling, punctuation, capitalization, paragraph indentation, etc.,	• Very few mistakes in spelling, punctuation, capitalization, paragraph indentation, etc			

A2 LEVEL ASSESSMENT SCALE FOR WRITTEN WORK

(Taken from http://www.coe.int/t/DG4/Portfolio/documents/Framework_EN.pdf

http://epep.at (go to: Assessment → Useful assessment scales)

APPENDIX - F: The Writing Task Assessment Rubric Used by the Teachers

	0	1	2	3	4	5
Content	• No relevance to the task or task not attempted • Little or no development of ideas	• Considerably irrelevant to the topic • Shows little knowledge of subject	• Inadequate coverage of the topic • Addresses the topic using inappropriate and/or insufficient ideas	• Adequate coverage of topic • Addresses the topic using somewhat appropriate and/or sufficient ideas	• Good coverage of topic • Addresses the topic using clear, appropriate, and sufficient ideas	• Very good coverage of topic • Fully developed passage with very good justification
Organization	• Completely disorganized • Lacks unity and coherence	• Disorganized (inappropriate format) • Mostly incoherent; ideas disconnected	• Inadequately organized (badly expressed topic sentence, repetitive and unclear concluding sentence, insufficient supporting ideas) • Considerably incoherent; ideas are not well connected	• Adequately organized (a weak topic sentence, unbalanced development of supporting ideas) • Mostly coherent and fluent passage; ideas flow meaningfully and logically, but may contain some redundancy and some unclear connections	• Generally well-organized (sufficiently developed topic sentence, supporting ideas, and concluding sentence) • Coherent and fluent passage; ideas generally flow meaningfully and logically	• Well-organized (a clear topic sentence, fully developed supporting ideas, and a good concluding sentence) • Very coherent and fluent passage
Vocabulary	• Very limited range • Very poor knowledge of words, idioms, and word forms	• Uses limited range of words, most of which are inaccurate or irrelevant • Translation-based errors	• Shows some use of varied word choice that is relevant to the content • Frequent word/idiom errors; inappropriate choice and usage	• Uses adequate range of vocabulary that is relevant to the content • Some errors of word/idiom choice, but effective transmission of meaning.	• Uses a wide range of vocabulary that is relevant to the content effectively and accurately • Very few mistakes in the word/ idiom choice	
Grammar	• Serious lack of language	• The language is rarely clear • Significant defects in the use of complex constructions; frequent errors in agreement, number, tense, negation, word order, articles, pronouns, prepositions	• Several problems with using appropriate grammatical structures, which affect comprehension • There are occasional errors in agreement, number, tense, negation, word order, articles, pronouns, prepositions	• No major difficulties in the use of appropriate grammatical structures • Some problems in the use of complex constructions; a few errors in agreement, number, tense, word order, articles, pronouns, prepositions	• Uses a wide range of level structures effectively and accurately • Very few mistakes agreement, number, tense, word order, articles, pronouns, prepositions	
Mechanics	• No control over spelling and punctuation, capitalization, and paragraph indentation etc,	• Occasional errors in spelling, punctuation, capitalization, paragraph indentation, etc., which do not interfere with meaning	• Masters conventions of spelling, punctuation, capitalization, paragraph indentation, etc			

The objective of this survey is to collect tangible information about your experiences and views on CEFR studies. All information provided by you will be kept confidential, and used merely for academic purposes.

PERSONAL DETAILS:

Name / Surname: **Gender:**

Age: **Years of Experience:**

1. Are you trained or informed about CEFR descriptors and standards? What do you know about the CEFR?

2. What are your viewpoints on the benefits of assessing writing in accordance with CEFR descriptors?

3. What are your viewpoints on the contribution of the CEFR to teacher development?

4. Do you think the criteria you use while assessing writing tasks are based on the CEFR descriptors?

5. How does the spotcheck done by the experts to standardise assessing written productions of the students help you? And to what extent do your gradings conform with the CEFR experts' gradings?

T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

2 2 Ocak 2018

Sayı : 35853172/433-316

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi: 11.01.2018 tarih ve 77 sayılı yazınız.

Enstitünüz Yabancı Diller Eğitimi Anabilim Dalı İngiliz Dili Eğitimi Bilim Dalı tezli yüksek lisans programı öğrencilerinden **Hamdiye AVCI**'nın **Prof. Dr. İsmail Hakkı MİRİCİ** danışmanlığında yürüttüğü **"Yazma Becerisinin Avrupa Dilleri Ortak Çerçeve Programı (CEPR) Tanımlayıcıları Doğrultusunda Değerlendirilmesi"** başlıklı tez çalışması, Üniversitemiz Senatosu Etik Komisyonunun **16 Ocak 2018** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini rica ederim.

Prof. Dr. Rahime M. NOHUTCU
Rektör a.
Rektör Yardımcısı

Hacettepe Üniversitesi Rektörlük 06100 Sıhhiye-Ankara
Telefon: 0 (312) 305 3001 - 3002 • Faks: 0 (312) 311 9992
E-posta: yazimd@hacettepe.edu.tr • www.hacettepe.edu.tr

Ayrıntılı Bilgi için:
Yazı İşleri Müdürlüğü
0 (312) 305 1008

152

APPENDIX - I: Declaration of Ethical Conduct

I hereby declare that…

- I have prepared this thesis in accordance with the thesis writing guidelines of the Graduate School of Educational Sciences of Hacettepe University;

- all information and documents in the thesis/dissertation have been obtained in accordance with academic regulations;

- all audio visual and written information and results have been presented in compliance with scientific and ethical standards;

- in case of using other people's work, related studies have been cited in accordance with scientific and ethical standards;

- all cited studies have been fully and decently referenced and included in the list of References;

- I did not do any distortion and/or manipulation on the data set,

- and **NO** part of this work was presented as a part of any other thesis study at this or any other university.

27/09/2019

Hamdiye AVCI

APPENDIX - J: Thesis/Dissertation Originality Report

23/08/2019

HACETTEPE UNIVERSITY

Graduate School of Educational Sciences

To The Department of Foreign Language Education

Thesis Title: Assessing Writing in Accordance with the Descriptors of the Common European Framework of Reference for Languages.

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defence	Similarity Index	Submission ID
25/04/2019	171	191.293	22/08/2019	%16	1118874876

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname:	Hamdiye AVCI
Student No.:	N13225490
Department:	Foreign Language Education
Program:	English Language Teaching
Status:	☒ Masters ☐ Ph.D. ☐ Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED

(Assist.Prof.Dr. Hatice ERGÜL)

APPENDIX - K: Yüksek Lisans Tez Çalışması Orijinallik Raporu

23/08/2019

HACETTEPE ÜNİVERSİTESİ

Eğitim Bilimleri Enstitüsü

Yabancı Diller Eğitimi Ana Bilim Dalı Başkanlığına,

Tez Başlığı :Yazma Becerisinin Avrupa Dilleri Öğretimi Ortak Çerçeve Programı (CEFR) Betimleyicileri Doğrultusunda Değerlendirilmesi.

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
25/04/2019	171	191.293	22/08/2019	%16	1118874876

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve 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 eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Hamdiye AVCI

Öğrenci No.: N13225490

Ana Bilim Dalı: Yabancı Diller Eğitimi

İmza

Programı: İngiliz Dili Eğitimi

Statüsü: ☒ Y.Lisans ☐ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

(Dr. Öğr. Üyesi Hatice ERGÜL)

155

APPENDIX - L: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan **"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına ilişkin Yönerge"** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- o Enstitü/Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. [1]

- o Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. [2]

- o Tezimle ilgili gizlilik kararı verilmiştir. [3]

27 /09 /2019

Hamdiye avcı

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

(1) *Madde 6. 1. Lisansüstü tezle ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü Üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.*

(2) *Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanın önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.*

(3) *Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü Üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.*
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

* Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.