POLICIES OF MINISTRY OF NATIONAL EDUCATION IN TERMS OF EVALUATION IN FOREIGN LANGUAGE EDUCATION

Ahmet ÖNAL [1] **Adnan YILMAZ**[2]

ABSTRACT

Evaluation, being one of the basic stages of education process, has a deterministic role on decisions taken as to both the stakeholders (students and teachers) and the other stages of education process. As the main aim of Foreign Language Education is to equip the learners with a skill rather than simply transfer pre-specified knowledge to them, evaluating to what extent this aim has been achieved brings with it certain difficulties. Within this article, current policies of Ministry of National Education as to primary, secondary and tertiary levels of education have been dealt with and solutions for the problems identified have been suggested.

Keywords: Foreign Language Education, Evaluation, Ministry of National Education, Foreign Language Education Policy.

[1] Okutman, Süleyman Demirel Üniversitesi, Yabancı Diller Yüksekokulu. ahmetonal32@gmail.com
[2] Araştırma Görevlisi, Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi. adnanyil@gmail.com

MİLLİ EĞİTİM BAKANLIĞI'NIN YABANCI DİL EĞİTİMİNDE

ÖLÇME-DEĞERLENDİRMEYE İLİŞKİN POLİTİKALARI

ÖZET

Ölçme değerlendirme, eğitim sürecinin temel aşamalarından biri olarak hem paydaş bireylerle (öğrenci ve öğretmen) ilgili verilen kararlarda hem de eğitim sürecinin diğer aşamalarına ilişkin verilen kararlarda belirleyici bir role sahiptir. Yabancı Dil eğitiminin esas amacı öğrenenlere bilgi kazandırmaktan ziyade bir yetenek kazandırmak olduğu için, bu amaca ne ölçüde ulaşıldığının değerlendirilmesi de çeşitli zorlukları beraberinde getirmektedir. Bu makalede Milli Eğitim Bakanlığı'nın ilk, orta ve yüksek öğrenim düzeylerinde Yabancı Dil Eğitimine ilişkin mevcut politikaları ele alınarak, eksiklikleriyle ilgili çözüm önerileri getirilmiştir.

Anahtar Kelimeler: Yabancı Dil Eğitimi, Ölçme-Değerlendirme, Milli Eğitim Bakanlığı, Yabancı Dil Eğitim politikaları.

1. Evaluation in Foreign Language Education

The term *evaluation* in language education, also in other educational domains, by and large "evokes images of an end-of-course paper-and-pencil test" prepared to inform both teachers and students about how much they have achieved in the process of teaching and learning in relation to the predetermined objectives (Coombe, Folse & Hubley, 2007: xiii). However, it is ascertained that evaluation goes far beyond the application of tests in the process of education (Demirel, 2010; Coombe et al., 2007).

Regarding the points above, the rationale behind making evaluations in language classrooms is considered to have a multi-fold characteristic (Coombe et al., 2007: xvi-xvii): (1) Teachers try to *place students* in the right level of classroom instruction via the assessments and evaluations they make. The purpose of such evaluations is to help students benefit from instruction at the right level. (2) Another reason for making evaluations is suggested to be *diagnosing student problems* (e.g., determining students' strengths and weaknesses). (3) Teachers also aim to *find out about the language proficiency of their students* and determine whether their students can meet the benchmarks prescribed. (4) Evaluation is also done to determine students' *academic performance* in order to make decisions about their achievements regarding the course goals and mastery of course content offered. With such a purpose in mind, learners either pass or fail the teaching program. (5) Making evaluations in education is also corned with *instructional decision-making*. Teachers try to assist their students in the most appropriate and effective way via deciding which material to present next and/or what to revise in order to meet the course objectives and students' immediate needs. In this respect, Chastain (1988) asserts that teachers need to evaluate constantly their teaching on the basis of student reaction, interest, motivation, preparation, participation, perseverance and achievement. Such evaluations can also lead to making updates in the curriculum offered to students.

With the points above in mind, the mode of assessment has gone through a dramatic change over the course of time. In this aspect, the underlying theoretical framework of instruction is posited as one of the primary cause for the shift in measures (e.g., quizzes, tests, projects, portfolios, and so forth) used to evaluate students' achievements. Such a shift is evident particularly in *constructivism* which is the leading theory forming the basis for the delivery of instruction in today's educational arena. Constructivism provides a wide theoretical framework from the design of learning settings to the construction of assessment processes. Therefore today, while there is a shift from traditional to student-centered learning settings, there are also innovations in assessment procedures, where the change is from summative assessment to formative assessment (Yurdabakan, 2011). The theoretical framework that emerged with constructivism has caused learning settings to be student-centered and have triggered the assessment processes to include questioning of learning process. As claimed by constructivist approach, "knowledge is formed when an individual interacts with his/her environment" (Yurdabakan, 2011: 52). Therefore, as Yurdabakan argues, "instructional practices should stress high level thinking and deep understanding, focus on real life problems, give importance to students' meta-cognitive knowledge and focus on improving their reflective skills by making use of different assessment practices" (2011: 52).

Regarding all those points, various types of tests are administered to serve for different purposes and they are used at different stages of the course to gather information about students. Main types of tests that are utilized in foreign language education will be dealt with in more detail in the next section of the article.

1.1.   Types of Tests

As aforementioned, tests are utilized for numerous reasons in education, referring to language education in the present study. The most common use of language tests is contented to identify students' strengths and weaknesses in relation to their abilities (Coombe et al., 2007). Coombe et al. (2007: xvi) also assert that "[I]nformation gleaned from tests also assists us in deciding who should be allowed to participate in a particular course or program area". Another significant use of tests is stated to be providing information about the effectiveness of programs of instruction. Regarding these points, Coombe et al. identify different types of tests (see Table 1).

Table 1. *Types of Tests (based on the classification by Coombe et al., 2007)*

	Types of Tests	Purpose
1	*Placement Tests*	- To assess students' level of language ability and place them in an appropriate course or class, - To create groups of learners who are homogenous in level.
2	*Aptitude Tests*	- To measure capacity or general ability to learn a foreign or second language.
3	*Diagnostic Tests*	- To identify students' strengths and weaknesses - To determine further course activities and provide students with remediation.
4	*Progress Tests*	- To measure the progress that students make toward defined course or program goals.
5	*Achievement Tests*	- To determine what students have learnt with regard to stated course outcomes during mid- and end-point of the semester or academic year.
6	*Proficiency Tests*	- To assess the overall language ability of students at varying levels. - To describe what students are capable of doing in a language.

As is obvious in the table, there are six types of tests and these tests have various uses and purposes in language education (Coombe et al., 2007). *Placement tests* are generally used to place students in an appropriate course or class with regard to their level of language ability. While some institutions base the test content either on a theory of general language proficiency or on learning objectives of the curriculum, others base the test content on the aspects of the syllabus taught (Alderson, Claphan, & Wall, 1995). As for the *aptitude test*, Brown (2004) states that they are used predominantly in determining to assist a person for special training based on language aptitude (e.g., the Modern Language Aptitude Test/MLAT; Pimsleur Language Aptitude Battery/PLAB). *Diagnostic tests* are administered to determine students' strengths and weaknesses so that they could be provided with further help. Harris and McCann (1994) assert that preparing diagnostic tests is rather difficult, and therefore placement tests assume a dual role of

placement and diagnosis. As for the *progress tests*, they are administered at different stages throughout a language course in order to decide what has been learned and what has not been. Compared to progress tests, achievement tests are usually administered at mid- and end-point of the semester or academic year with a purpose to determine whether students have passed or failed a course or class. The primary distinction between these both tests, namely progress and achievement tests lies in their scopes and the decision made at the end of the tests. In the former, the focus is generally narrower because it covers less material and assesses fewer objectives. However, in the latter, the scope much larger since it covers materials from an entire course or semester, hence measuring more objectives. The last type of test in the table is identified as proficiency tests. *Proficiency tests* are not based on a particular curriculum or language program. They tell us how capable a person is in a certain language skill (e.g., speaking). These tests are usually prepared by external institutions such as Educational Testing Services (ETS), or Cambridge ESOL. TOEFL® and IELTS™ are some very good examples of this type of tests.

1.2. Additional Ways of Classifying Tests

Tests can also be classified in different ways regarding a number of points like scoring manner, determining a criterion, making decisions, impacting test takers, and so forth (Coombe et al., 2007). Regarding these points, Table 2 presents the additional ways of labeling tests and their purposes.

Table 2. *Additional Ways of Classifying Tests (based on the classification of Coombe et al., 2007)*

1	**Subjective** Scoring students' responses by the opinions or personal judgments of the rater	V E R S U S	**Objective** Scoring students' responses by comparing them with an established set of acceptable responses on an answer key
2	**Criterion-Referenced** Scoring students' performance by comparing only to the amount or percentage of material learned in the light of a criterion or cut-off score set in advance		**Norm-Referenced/Standardized** Interpreting students' scores relative to all other students who take the exam to spread students out along a continuum of scores
3	**Formative** Improving instruction and providing feedback to students		**Summative** Determining whether students can move on to a higher level
4	**Low-Stakes** Having minor impact on the lives of individuals or on small programs (e.g., in-class progress tests, short quizzes)		**High-Stakes** Having a major impact on the lives of large numbers of individuals or on large programs (e.g., TOEFL)

As can be seen in Table 2, tests are classified as subjective and objective depending on the manner of the rater. In objective tests, the scorer is not entailed to have particular knowledge or training the area concerned; however, in the subjective test, the rater is required to have prior knowledge and training in the area concerned. Test formats including multiple-choice questions, True/False/Not Given, and matching activities are associated with objective tests, whereas subjective tests are associated with essay tests, interviews, or comprehension questions. As for criterion-referenced versus norm-referenced tests, they differ from each other in a number of ways. Criterion-referenced test (CRTs) are usually designed to assess mastery of predetermined instructional objectives pertaining to a specific course or program, while norm-references tests (NRTs) are developed to measure global language abilities (Brown, 2005). In CRTs, students' scores are interpreted according to the criterion or cut-off score determined in advance; in contrast, in NRTs, students' scores are interpreted relative to all other students who take the exam. Regarding the decision made during or at the end of a course or class in relation to students' scores, summative and formative tests are administered. Summative tests are administered at the end of a course and used for deciding which students move on to a higher level (Harris & McCann, 1994). On the other hand, formative tests are carried out with the aim of "evaluating students in the process of 'forming' their competencies and skills with the goal of helping them to continue that growth process" (Brown, 2004: 6). The last classification in this part is concerned with the impact of the tests on test-takers, namely high-stakes versus low-stakes tests. High-stakes tests have substantial effect on the lives of large number of individuals or on large programs. For instance, the TOEFL® test is high stakes since it is used for admission into a university program, hence having a significant influence on test-takers. However, those tests (e.g., in-class progress test, short quizzes, and so on) which have minor effect on the lives of the individuals or on small programs are categorized as low-stakes tests.

1.3. Traditional versus Alternative Assessment

A comparison of alternative assessment with traditional one yields a number of points that may help distinguish them from each other. Alternative assessment requires the students to show what they can do, integrate and produce rather than recall and reproduce (Huerta-Macias, 1995). In the light of this overall explanation, several types of alternative assessment are suggested in today's language classrooms (MoNE, 2006; Coombe et al., 2007):

- writing assessment   - portfolio assessment
- classroom assessment   - self assessment
- teacher observation   - anecdotal records
- checklists   - rating scales
- scoring rubrics   - benchmark standards
- student-designed tests   - projects
- presentations

Considering the types of alternative assessments suggested above, it is noted that alternative assessment differs from traditional testing in that it (MoNE, 2006, p. 26; Coombe et al., 2007: xix; Heurta-Macias, 1995):

- does not intrude on ordinary classroom activities.
- provides information on each student's strengths and weaknesses.
- enables genuine implementation of the curriculum in the classroom.
- provides various indications that can be used to gauge student progress.
- has a performance-based and realistic nature.
- is criterion-referenced rather than norm-referenced.
- is inseparably integrated with teaching.
- spares room for students' creativity.
- measures the extent to which students can create, reflect, solve problems, collect and use information, and formulate interesting and worthwhile questions.
- considers individual differences.

Language performance is not restricted to merely having a command of the target language on basis of *form (structure)*. It rather goes beyond the form dimension into *meaning (semantic)* and *use (pragmatic)* dimensions, and therefore alternative assessment enables students to practice the target language in much more meaningful contexts and build pragmatic knowledge of it via authentic tasks (Larsen-Freeman, 2001). Of note at this point, alternative assessment does also have some shortcomings along with benefits. Self-assessment and portfolios serve to solidify this dichotomy effectively. These two types of alternative assessment are noted to have the following benefits (Yurdabakan, 2011: 61-62):

- Since individuals are at a more advantageous position than an outside observer to detect the changes about themselves, self-assessment could help the definition of those changes more realistically.
- Self-assessment improves students' self-criticism skills, while it increases their level of perception of their weak and strong aspects.
- Reflection and self-evaluation activities conducted through portfolio improve learners' meta-cognitive skills.
- By encouraging creativity, diversity and autonomy, it fosters learners' higher level thinking skills.

On the other side of the dichotomy, Yurdabakan (2011) underscores some drawbacks, and these are:

- Portfolio assessment is an expensive and long-term evaluation tool.
- One of the widespread opinions about self-assessment is the concern over students' giving higher scores to themselves.

Departing from this information provided on evaluation, assessment, and testing, various tests and ways of assessments are employed at primary and secondary schools in Turkey by the Ministry of National Education (MoNE) in order to evaluate students' performance in foreign language learning. In the following parts, these points will be explicated.

1.4. The Cornerstones of Testing

Sarıçoban (2011) examined the current situation in test (a) construction: designing, structuring, developing, (b) administering, and (c) assessing the foreign language tests by collecting samples of oral and written tests of 5 English Language teachers working at a state high school in Ankara and has given some suggestions on this issue. In his analysis of the documents (tests/exams) gathered, the researcher employs fourteen criteria:

Validity: It is "…the extent to which it (*the test*) measures what it is supposed to measure and nothing else" (Heaton, 1990: 159). Four types of validity are explained by Heaton; namely, face validity, content validity, construct validity, and empirical/statistical validity.

Reliability: A reliable test "… produces essentially the same results consistently on different occasions when the conditions of the test remain the same" (Madsen, 1983: 179).

Backwash effect: '*Backwash*' (also known as '*Washback*') is known as the effect of testing on teaching and learning and can be harmful or beneficial. "If a test is regarded as important, if the stakes are high, preparation for it can come to dominate all teaching and learning activities. And if the test content and testing techniques are at variance with the objectives of the course, there is likely to be harmful backwash" (Hughes, 2003: 1).

Language skills and areas: In FLT four language skills (Listening, Speaking, Reading, and Writing) as well as language areas (Pronunciation, Vocabulary, Grammar, and Translation) are aimed to be improved. As noted above, the curriculum reform in FLE by MoNE aims to promote learners' communicative proficiency in English and oral skills of the students should also be included into the evaluation process as well as pronunciation.

Contextualization: Contextualization is the process of "…placing the target language in a realistic setting, so as to be meaningful to the student" (Sarıçoban, 2011: 400). The teacher should always try to write items as similar as possible to language that would occur in actual usage. Testing in context implies that often a contextual situation must be established to clarify for the students exactly what the correct answer would be (Chastain, 1988).

Time: The duration of the test should be stated on the paper and the teacher should allow sufficient time to finish the test.

Typing and Layout: It is advisable to use the computer to type the tests rather than handwriting as the handwriting may be difficult to read and also the layout of the test paper should be designed clearly.

Language proficiency (simple or complex structures): As Sarıçoban (2011: 400) notes, "The language used in the test must be appropriate to the proficiency level of the learners and the language level of learners should be taken into

account when constructing the test." Moreover, the teacher should make sure that the items have been sequenced from easy to more difficult on the test.

Instructions: The instruction(s) given on the test must be clear and understandable for the test-takers.

Motivation: The teacher should pay special attention to the design of the items in order not to cause anxiety among the test-takers.

Scoring: Considering the high work-load of the teachers in Turkey, the teacher should make the test as easy as possible to score and objectiveness should also be taken into consideration in the scoring process.

Spelling: The teacher may ask his/her colleagues to proofread the test prior to administrating it.

Diagnostic testing: A diagnostic test is designed in order to diagnose or identify pre-specified aspects of a language skill. "Diagnostic tests are used to identify learners' strengths and weaknesses. They are intended primarily to ascertain what learning still needs to take place" (Hughes, 2003: 15). The teacher may design his/her later instruction considering the weaknesses of the test-takers.

Homework: There should be little difference between the normal classroom activities and the test items (Chastain, 1988). The teacher should try to acquaint the students with the techniques by providing them with handouts and worksheets.

Sarıçoban (2011: 401-402) concludes that the sample tests he has examined lack validity as they do not evaluate writing and listening skills of the students although they appear in the course book. Tests are reported have a high level of reliability; however, they have a harmful washback effect in that they do not evaluate students' oral skills (listening and speaking) although the course book has an accompanying CD to improve their listening skills. Sarıçoban (2011: 403) claims that the sample tests do not have any problems in terms of contextualization, time, typing, instructions, scoring and spelling and punctuation. Although not exhaustive, these fourteen criteria put forward by Sarıçoban (2011) should be taken into consideration by FL teachers in the process of preparing tests. Moreover, these criteria may also be used to examine the tests administered by FL teachers.

2. Policies on Foreign Language Evaluation in Turkey

The Ministry of National Education (MoNE) and Student Selection and Placement Center (OSYM) are the two main bodies responsible for administering tests in the Republic of Turkey. These two bodies of test administration carry out a number of language assessments for a variety of purposes.

Student Selection and Placement Center (OSYM), by and large, administer high-stakes tests which enable the test-takers to get a job, a diploma, a scholarship or a license to practice a profession, thus bearing a crucial impact of the lives of test-takers. Turkish educational system offers many different high-stakes tests to the students who are constantly preparing for a specific examination to possess a better education at different levels of education. KPDS,

UDS, and YDS that claim to evaluate test-takers' FL proficiency are some very good examples of high-stakes tests administered in Turkey. In general, these tests assess test-takers' reading skills, vocabulary knowledge, and knowledge of grammar as well as their translation skills. Regarding these points, these proficiency tests are often criticized as they lack validity and reliability as well as their negative washback effects. For instance, Özmen (2011) studied the washback effects of SEPPPO (KPSS) on prospective English language teachers and concluded that SEPPPO exerts negative and harmful effects not only on these student-teachers but also on educational faculties and families.

In Turkey, the main rules to be obeyed by Foreign Language (FL) teachers in their conduct of evaluation at primary and secondary schools are issued by the MoNE, Board of Education and Discipline (BoED) in the form of written 'Regulations'. However, FL teachers are given a limited freedom as to how they will carry out the evaluation. They are required to decide on the details of the process of evaluation during their FL Unit Teachers' Meetings. The lack of regular supervision and in-service trainings lead to lack of unity in the conduct of evaluation across the institutions. To tackle the problems with the methods of assessment in the process of language education, the suggested "English Language Curriculum for Primary Education (grades 4, 5, 6, 7 and 8)" by the Board of Education and Discipline (2006) recommends the use *alternative* assessment methods by referring to the '*affective filter theory*' in that the *traditional* tests may cause anxiety in test-takers and therefore they are detrimental to the learning process. In the light of these explanations, all nonconventional ways of assessment are tagged as "*alternative* assessment" or "*authentic* assessment". The MoNE (2006) posits the following types of assessment as authentic (see also Part 1.3. Traditional versus Alternative Assessment):

- writing assessment   - portfolio assessment

- classroom assessment   - self assessment

- teacher observation   - anecdotal records

- checklists   - rating scales

- scoring rubrics   - benchmark standards

Of these authentic types of assessments, portfolio assessment has gained great significance as a way of evaluation. In this respect, Sarıçoban (2011: 406) states that:

"The portfolio enables the language learners to keep record of their language learning and cultural experiences either at school or outside school. The 'can do' descriptors of the CEFRL [Common European Framework of Reference for Languages] is fundamental to ELP [European Language Portfolio] as without them the language learner would not be able to keep track of his/her own progress in a detailed and constructed way as it is possible with the CEFRL."

Sarıçoban also notes that ELP consists of three parts, and these are:

1. *Language Passport:* It provides an overview of the individual's proficiency in different languages.

2. *Language Biography:* It encourages the language learner to record his/her personal development of language/s, his/her learning process, what he/she can do in each language, linguistic and cultural experiences gained in and/or outside formal educational contexts, and

3. *Dossier:* It enables the language learner to choose materials to report and to document achievements.

Sarıçoban further maintains that the recognition and implementation of such an alternative way of evaluation may replace traditional tests, and a more rational and acceptable way of assessing foreign language learners' achievements could be possible.

2.1. Current Regulations on Foreign Language Evaluation by MoNE

Kırkgöz (2007b: 180-181) notes that "Assessment is an important part of the curriculum and the teacher is the only assessor" according to the regulations issued by the MoNE. In relation to this notion, Kırkgöz and Ağçam (2012) add that two important curriculum reforms in English language teaching have taken place since the late 1990s, the former in 1997 and the latter in 2005. With the reform in 1997, the MoNE required the application of innovative approaches, namely Communicative Language Teaching (CLT), in foreign language instruction in primary and secondary education. Moreover, the teaching of English was lengthened to grades 4 and 5, and assessment became an integral part of the CLT.

With this reform, students at 4th and 5th grades were required to have two written tests per semester. The tests and written examinations are to be done in an indirect way and are entailed to evaluate curriculum objectives. It is also stated that students' performances should be evaluated via "written, spoken, and practical examinations, homework assignments, and projects by the teacher…" (Kırkgöz & Ağçam, 2012: 123). As for the reforms in 2005, the MoNE suggested alternative ways of assessing language learners' performances, including portfolios, peer assessment, self-assessment. The rationale behind these alternative ways of assessment is "to complement the formal assessment practices with less quantitative ways" (Kırkgöz & Ağçam, 2012: 123). In the light of these innovations, the assessment and evaluation regulations at primary, secondary and tertiary schools will be dealt with in the following parts.

2.2. Evaluation of Foreign Language Education at Primary Schools

The general guidelines used for evaluating students' performance at primary schools are determined by the MoNE as follows (MoNE, 2012):

- The educational year consists of two semesters that complement each other in terms of assessment and evaluation.

- The well-defined objectives and learning outcomes are taken into consideration while measuring and evaluating student achievements.

- In testing and evaluation, unity is warranted across schools, provinces and the whole country.
- Students' success is determined according to the tests administered, projects and other studies that students have been assigned.
- Students are obliged to participate in lessons, and applications and evaluation activities.
- The activities related to determining students' success consist of participation in lessons and activities and performance-oriented duties.
- Students' cognitive, affective, social and psychomotor characteristics should be considered as a whole while evaluating students' performances.
- Tasks and activities that assess such points as critical-thinking, creativity, research, investigation, problem-solving, and so forth should be focused on in testing tools and methods.
- The assessment tools need to be reliable, valid and practical.

The interpretation of the scores is carried out according to the cut points specified in Table 3. As can be seen in the table, the minimum passing grade is 2 out of 5 (the top grade):

Table 3. *Criterion for Evaluation of Test Scores (MoNE, 2006)*

Score	Grade	Degree
85-100	5	Excellent
70-84	4	Good
55-69	3	Middle
45-54	2	Passing
25-44	1	Failing
0-24	0	Ineffective

Students at grades 1, 2 and 3 are not given any exams (MoNE, 2012). Their performance is evaluated via the course teacher's observations and the projects and performance-oriented activities. In 4th grade and above, students' performances are evaluated through exams, projects, and other performance-oriented tasks. As for the number of exams administered per lesson in each semester, the teachers need to give students at least 2 exams for the courses that are 3 hours or fewer a week and at least 3 exams for the courses that are more than 3 hours per week. Students are also assigned to prepare a project individually or in groups through the guidance of the course teacher. Students get a grade for a course per semester and the final grade is given according to the arithmetic mean of the both semesters.

MoNE (2012) further adds that the exams administered at schools may be designed in different formats, including (a) essay questions, (b) multiple-choice questions, matching format, short-answer items, open-ended questions, True/False format, cloze/gap-fill items, and so forth. In the first format, there should be at least three questions; however, in the latter, there should be multiple questions with short answers. Course teachers are also required to prepare an answer key for each exam to be administered. Moreover, teachers are entailed to prepare and administer at least one of the exams together in order to augment the cooperation among teachers and facilitate better evaluation.

2.3. Evaluation of Foreign Language Education at Secondary Schools

The general guidelines stipulated by the "Regulation of Passing and Exams for Secondary Schools" by MoNE (2010) show great similarities to those used at primary schools. The guidelines are as follows:

- The curriculum is taken into consideration in the implementation of evaluation.
- The assessment tools need to be reliable, valid and practical.
- The achievement of the students is continuously observed by several methods and techniques.
- The result of each assessment procedure is linked to the instruction provided in line with the capacity and ability of the students.
- The number of written and oral exams to be conducted throughout the year, the assessment of homework given to the students, the methods and the techniques to be employed during the instruction and similar issues are specified at the meeting of "FL Unit Teachers" conducted at the beginning of the academic year.
- The students' grades are determined by taking into account their performance on written, spoken, and practical examinations, homework assignments, and projects.
- The results of the assessment are used to determine to what extent the goals of the instruction have been met, to identify the parts of the curriculum that the students have been successful and unsuccessful on and to detect the parts of the curriculum that need remedial teaching.
- The number of written examinations should be at least 2 for the courses that are 1 or 2 hours per week whereas they should be at least 3 for the courses that are 3 or more hours a week.
- The number of questions on an essay type exam must be at least 5 whereas in multiple-choice type, matching type, open ended type and true/false type of exams, the number of questions must be more.
- Each term, the teachers need to give an oral exam grade to the students and in accordance with the features and intensity of the course, for the courses that are (more than) 3 hours a week, a maximum of 3 oral exam grades, and for the courses that are 1 or 2 hours a week, a maximum of 2 oral exam grades are to be given.

The interpretation of the scores that students get is also done according to the figures presented in Table 3. That is, the cut point to move on to higher grade is determined to be 2 out of 5 (the top grade). Another important feature of the exams employed is that teachers are entailed to prepare and administer at least one of the exams together in order to augment the cooperation among teachers and facilitate better evaluation, hence showing similarity to the examination guidelines pursued at primary schools.

The MoNE also provides the FL teachers with a sample project assignment and its rubric in the English Language Curriculum for Secondary Education issued by Board of Education and Discipline (see Tables 4 and 5):

Table 4. *Sample Performance Assignment (MoNE, 2011: 35)*

Content Level	English
Language Proficiency Level	B1
Theme	Science and Technology
Topic	Environmental Consciousness
Expected Performance	Research, Writing Skills
Duration	2 Weeks
Grading	Scoring Rubric

Topic of Assignment

Dear Students,

You are required to conduct a research into the works of volunteer environmental organizations in Turkey and present your findings in the form of a report by comparing the similar and different works of these organizations. While carrying out this assignment, you should include;

- The preventive measures taken by these organizations to protect the environment,

- The organizations, institutions and individuals that collaborate with these organizations,

- The extent of their success in their attempts to protect the environment.

Pay attention to the following while carrying out the assignment

1- You should make a plan so as to complete the assignment in time.

2- You can check the addresses of volunteer environmental organizations on the internet.

3- You can compile the necessary information from the printed or visual media, journals, books, brochures or (if possible) people who are members of these organizations.

4- If you cannot obtain satisfactory answers for the questions in this project assignment, you can ask your questions to the related organizations by sending an e-mail.

5- The information you will compare should not be more than a few pages.

6- Make sure you use a clear and understandable language in your report.

7- The sources that have been used must appear in the 'references' section.

8- You should enrich your report by integrating pictures, photographs, drawings and graphs related to the works

of the organization.

9- Make sure you display your report for your friends.

Scoring Rubric has been prepared to inform you about the issues that will be taken into consideration while evaluating your assignment. Please do not write anything on it as it will be completed by your teacher.

Retrieved on March 15, 2013 from http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72

The sample performance assignment states that the 'language proficiency level' of the students is B1, which does not seem to be the real case. The theme of the assignment is 'Science and Technology' and probably it will be an interesting theme for students who attend 'Science' classes. Expected performance by the assignment is 'research and writing skills' and this again highlights the fact that oral skills of the students are, most of the time, neglected. The duration of the assignment is stated as 2 weeks and for such a comprehensive research, 2 weeks does not seem to be appropriate. The assignment is given for individual students but it could be a better idea to form groups of 3 or 4 students and ask them to hand in a single paper by collaborating with each other. The students are asked to list the sources they have used in the 'references' section of the paper. However, it is doubtful whether they have been taught such formal writing skills and procedures in their English classes. It would be unfair to ask the students to do something without first teaching it. Another point to consider is that the page limit has not been stated clearly and this uncertainty may lead to problems.

Table 5. *Scoring Rubric for the Assessment of the Project (MoNE, 2011: 36)*

	5	4	3	2	1
FORM					
The report has been designed well enough to manifest the similar and different points.					
The report has been enriched by pictures, photographs, drawings and graphs.					
The sources used in the report have been stated.					
CONTENT					
The preventive measures taken by these organizations to protect the environment have been included.					
The organizations, institutions and individuals that collaborate with these organizations are specified.					

The similarities and differences of the organizations and institutions have been explained.					
Comparisons of similarities and differences as to the extent of their success in their works to protect the environment have been done.					
LANGUAGE and WORDING					
A clear, fluent and understandable language has been used in the report.					
TOTAL					
TOTAL SCORE					

The maximum score for this rubric is 40 and the minimum score is 8. By using the formula below, this grade can be converted into the grading system out of 100.

Example: Suppose that the student got 30 according to the rubric.

$$\frac{40}{100} \qquad \frac{30}{x}$$

$$x=75$$

Retrieved on March 15, 2013 from http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72

The scoring rubric consists of 3 sections; *form* (15/40 – 37,5 %), *content* (20/40 – 50 %), and *language and wording* (5/40 – 12,5 %). The calculation of the total score may be difficult and confusing for the teacher because it is out of 40 and a formula is given to convert the total score into the grading system out of 100. The 'form' here refers to the form of the paper and not to the form of the language. The 'content' section is devoted to the data included in the paper. Only the 'language and wording' section of the rubric is about the foreign language skills of the students and its weight is just 12,5 %. In fact, this rubric does not evaluate the writing skills or foreign language skills of the students. Sarıçoban (2001: 114) suggests the following points to take into consideration in an effective writing process:

1. Purpose: the reason for writing
2. The Audience: the readers
3. Grammar: rules for verbs, subject-verb agreement, prepositions, article system, etc.
4. Syntax: sentence structure, sentence boundaries, word order, stylistic choices, etc.
5. Mechanics: spelling, punctuation, etc.
6. Organization: paragraphs, cohesion and unity, etc.
7. Word Choice: vocabulary

8. Content: relevance, clarity, originality, logic, etc.

9. The Writer's Process: getting ideas, getting started, writing drafts, revising, and final draft.

Most of these points are lacking in the rubric for the project assignment which is claimed to evaluate the writing skills of the students, and therefore, the rubric is far from being an effective evaluation tool.

2.4. Current Applications of Foreign Language Evaluation within Tertiary Level

Most of the universities in Turkey offer their students Preparatory Classes in their first year and the students are given an exemption exam before they begin Preparatory Class. Maden *et al.* (no date) studied the samples of exemption exams of 31 different universities in Turkey and concluded that:

- the implementation of Prep. Classes and exemption exams exhibits great diversity due to the fact that each institution has a different and unique regulation.
- the passing grade is 60 out of 100 in some institutions whereas it is 70 in others and even 80 in some others.
- in most of the institutions' exams (29 out of 31), listening and speaking skills of the students are not tested.
- the proportion of 4 skills in these exams are far from satisfactory.
- most of the exams investigated consisted of multiple-choice type of questions and thus lacking in testing the productive skills of the test-takers.
- some institutions ask their students to write a paragraph on a given topic, while some others require their students to write essays. However, most of the institutions do not test their students' writing skills.
- most of the institutions do not announce the exam results in accordance with the Language Proficiency Levels (A1, A2, B1, B2, C1, C2) specified by CEFR.

3. Some Final Remarks on Foreign Language Evaluation within MoNE

Kırkgöz (2007a) argues that the curriculum reforms in FLE by MoNE aim to promote learners' communicative proficiency in English, and these reforms have also brought about innovations at the level of testing/assessment. Instead of traditional 'paper and pencil' tests, performance-based assessment is proposed through implementing portfolios, which is claimed to be more congruent with the principles of the communicative language teaching. Portfolio assessment, an example of authentic assessment, focuses on documenting a student's language progress and performance. It enables teachers and parents to discuss and review the child's development on a concrete basis and is intended to complement the traditional product-oriented Turkish assessment system. Kırkgöz (2007: 225) further maintains that:

"The revised curriculum now calls for more performance-based assessments that align with the current views of curriculum development, more accurately reflecting children's language acquisition process. Overall, the suggested evaluation device in the recent curriculum document is based on the European

Language Portfolio. Along with this proposal, the new curriculum also highlights the necessity of training teachers in effectively implementing this new assessment."

It has been observed by Sarıçoban (2011) that in Turkey, assessment of productive skills of the FL learners are often neglected by most of the teachers working for the MoNE. Sarıçoban (2011: 398) notes that:

I recall the first two years of my undergraduate study in the department of English Language Teaching (ELT) where we used to be given pen-paper tests in our "Speaking" course midterms. It is still odd as it was years ago. For instance, in "speaking" courses at the ELT departments students are still asked to perform in a pen-paper test for their midterms; whereas, they should be tested orally. Dialogue completion tests and/or discussion type of tests on a topic (in written form for speaking) usually favored by the teacher are fashionable for use today. Of course, the lecturers of this speaking course seem to claim that they do not have enough time to administer a speaking test in their midterms since the classes are overcrowded. This is the same case in respect to the state high schools in our country.

4.   Conclusion and Suggestions

The success of FL Education in Turkey is clearly far from satisfactory and the FL Policies in Turkey have often been criticized by professionals and scholars (Sarıçoban, 2011; Özmen, 2011; Tılfarlıoğlu & Öztürk, 2007). Many suggestions aiming to improve the quality of FL evaluation in Turkey have been made by Daloğlu & Seferoğlu (2009: 30-33) in that FL teachers should be able to:

- Demonstrate an understanding of the purposes of assessment and use results appropriately,
- Demonstrate an understanding of the quality indicators of assessment instruments,
- Demonstrate an understanding of the limitations of assessment situations and make accommodations for students,
- Understand, develop, and use criterion-referenced assessments appropriately with learners,
- Understand, construct, and use assessment measures for a variety of purposes for students,
- Assess learners' language skills and communicative competence using multiple sources of information,
- Use performance-based assessment tools and tasks that measure learners' progress toward curriculum goals,
- Prepare students to use self- and peer-assessment techniques when appropriate.

In addition to these recommendations, testing should be parallel to the task done in the class. In other words questions should be contextualized and must be meaningful (Tılfarlıoğlu & Öztürk 2007). Sarıçoban (2011) recommends the recognition and application of ELP as a more rational and reasonable way of assessing foreign language learners' success. In-service training courses with the aim of improving FL teachers' evaluation skills and informing them of the developments in ELT need to be organized.

REFERENCES

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

Board of Education and Discipline, MoNE. (2006). *English Language Curriculum for Primary Education (grades 4, 5, 6, 7 and 8).* Retrieved from http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72 on March 8, 2013.

_____. (2011) English Language Curriculum for Secondary Education. Retrieved from http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72. on March 8, 2013.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains: NY: Longman.

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment.* New York: McGraw Hill.

Chastain, K. (1988). *Developing Second-Language Skills: Theory and Practice.* Harcourt Brace Jovanovich Publishers: USA.

Coombe, C., Folse, K., & Hubley, N. (2007). *A practical guide to assessing English language learners.* Ann Arbor, Michigan: The University of Michigan Press.

Daloğlu, A., & Seferoğlu, G. (2009) *Özel Alan Yeterlikleri İngilizce Komisyonu II. Dönem Raporu.* Retrieved March 17, 2013 from http://otmg.meb.gov.tr/belgeler/raporlar/ingilizce%20rapor%202.pdf.

Demirel, Ö. (2010). *Yabancı dil öğretimi.* Ankara, Turkey: Pegem Yayıncılık.

Harris, M., & McCann, P. (1994). *Assessment.* Oxford: MacMillan Heinemann.

Heaton, J. B. (1990). *Writing English language tests* (New Edition). Longman Group UK Limited.

Heurta-Macias, A. (1995). Alternative assessment: answers to commonly asked questions. *TESOL Journal,* 5, 8-10.

Hughes, A. (2003). *Testing for language teachers*. (Second Edition). UK: CUP

Kırkgöz, Y. (2007a) English language teaching in Turkey: Policy Changes and their Implementations. *RELC Journal,* 2007 38(2), pp. 216-228.

_____. (2007b): Language planning and implementation in Turkish primary schools. *Current Issues in Language Planning,* 8(2), 174-191.

Kırkgöz, Y., & Ağçam, R. (2012). Investigating the written assessment practices of Turkish teachers of English at primary education. *The Journal of Language and Linguistic Studies,* 8(2), 119-136.

Larsen-Freeman, D. 2001. Teaching grammar. In M Celce-Murcia (Ed.), *Teaching English as a second or foreign language.* (pp. 251-266). Boston: Heinle and Heinle.

Maden, S. S., Ere, S. & Yiğit, C. (no date) *Yükseköğretim hazırlık sınıfı yabancı dil seviye tespit (yeterlik) sınavlarının karşılaştırılması ve Avrupa Ortak Başvuru Metni (CEF) ölçütlerine uygunluklarının araştırılması.* Retrieved March 6, 2013 from http://www.eab.org.tr/eab/oc/egtconf/pdfkitap/pdf/100.pdf.

Madsen, H. S. (1983). *Techniques in testing*. UK: OUP.

Ministry of National Education (MoNE) (2006). *Orta öğretim kurumlari sinif geçme ve sinav yönetmeliği.* Retrieved on March 10, 2013 from http://mevzuat.meb.gov.tr/html/25664_0.html

_____. (2010). *Millî Eğitim Bakanliği orta öğretim kurumlari sinif geçme ve sinav yönetmeliği.* Retrieved March 10, 2013 from http://mevzuat.meb.gov.tr/html/25664_0.html.

_____. (2011). *İngilizce dersi (4-8.Sınıflar) öğretim programı ile seçmeli İngilizce dersi öğretim programı.* Retrieved March 17, 2013 from http://ttkb.meb.gov.tr/program2.aspx/program2.aspx?islem=1&kno=31

_____. (2012). Milli Eğitim Bakanlığı ilköğretim kurumları yönetmeliği. Retrieved March 22, 2013 from http://mevzuat.meb.gov.tr/html/225_0.html.

Özmen, K. S. (2011). Analyzing washback effect of SEPPPO on prospective English teachers. *The Journal of Language and Linguistic Studies,* 7(2), 24-52.

Sarıçoban, A. (2001). *The Teaching of Language Skills*. Hacettepe-Taş Publishing: Ankara.

_____. (2011). A Study on the English Language Teachers' Preparation of Tests. *H. Ü. Eğitim Fakültesi Dergisi (H. U. Journal of Education),* 41 (2011), 398-410.

Tılfarlıoğlu, F., Y., & Öztürk, A., R. (2007). An Analysis of ELT Teachers' Perceptions of Some Problems Concerning the Implementation of English Language Teaching Curricula in Elementary Schools. *Journal of Language and Linguistic Studies,* 3(1).

Yurdabakan, İ. (2011) The View of Constructivist Theory on Assessment: Alternative Assessment Methods in Education. *Ankara University, Journal of Faculty of Educational Sciences,* 44(1), 51-77.

APPENDICES

Appendix 1: Sample Project Assignment

EK 2

Performans Görevi

İçerik Düzeyi	İngilizce
Dil Yeterlik Düzeyi	B1
Tema	Bilim ve Teknoloji
Konu	Çevre Bilinci
Beklenen Performans	Araştırma, yazma becerisi
Süre	2 hafta
Puanlama Yöntemi	Dereceli puanlama anahtarı

Ödev Konusu

Sevgili öğrenciler,

Sizden, ülkemizdeki gönüllü çevre kuruluşlarının çalışmaları ile ilgili bir araştırma yapmanız, bu kuruluşların benzer ve farklı çalışmalarını karşılaştırarak bir rapor hâlinde sunmanız istenmektedir. Bu ödevi hazırlarken;

- Gönüllü çevre kuruluşlarının çevreye verilen zararı engellemek için yaptıkları caydırıcı çalışmalar,
- Bu kuruluşların hangi kurum, kuruluş ve kişilerle iş birliği yaptıkları,
- Çevreye zarar verilmemesi için bu kuruluşların yaptıkları caydırıcı çalışmalarda ne ölçüde başarılı oldukları hakkında yazınız.

Ödevi Hazırlarken Dikkat Etmeniz Gerekenler

1) Araştırmayı zamanında tamamlayacak şekilde bir plan oluşturmalısınız.
2) Ülkemizdeki gönüllü çevre kuruluşlarının adreslerini öğrenmek için internetten yararlanabilirsiniz.
3) Araştırmanızı yaparken basılı ve görsel medya, dergiler, çeşitli kitaplar, broşürlerden veya çevrenizde bu kuruluşlara üye kişiler varsa onlarla görüşerek bilgi toplayabilirsiniz.
4) Ödev konusunda yer alan soruların cevapları için yeterli bilgi edinemezseniz, e-posta aracılığıyla ilgili kuruluşlara bu soruları yöneltebilirsiniz.
5) Karşılaştırma yapacağınız bilgiler birkaç sayfayı geçmemelidir.
6) Raporunuzda açık ve anlaşılır bir anlatım kullanmaya özen göstermelisiniz.
7) Raporunuzu oluşturmak için yararlandığınız kaynaklar, "kaynakça" bölümünde yer almalıdır.
8) Hazırlayacağınız raporu; ülkemizdeki gönüllü çevre kuruluşlarının çalışmalarına ait resim, fotoğraf, çizim, karikatür vb. ile zenginleştirmelisiniz.
9) Hazırladığınız raporu arkadaşlarınıza sergileyiniz.

Dereceli puanlama anahtarı ödevinizin hangi ölçütlere göre değerlendirileceği konusunda sizlere bilgi vermek için hazırlanmıştır. Bu puanlama anahtarınız öğretmeniniz tarafından doldurulacağından üzerinde herhangi bir işaretleme yapmayınız.

Sample Project Assignment (Board of Education and Discipline, MoNE, 2011: 35) http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72

Appendix 2: Sample Rubric for the assessment of the Project.

DERECELİ PUANLAMA ANAHTARI

	5	4	3	2	1
BİÇİM					
Rapor, benzer ve farklı yönleri ortaya koyacak şekilde doğru tasarlanmış.					
Rapor; resim, çizim, fotoğraf, karikatür vb. ile zenginleştirilmiştir.					
Sunulan raporda faydalanılan kaynaklar belirtilmiştir.					
İÇERİK					
Gönüllü çevre kuruluşlarının çevreye verilen zararı engellemek için yaptıkları caydırıcı çalışmalar hakkında bilgi verilmiştir.					
Bu kuruluşların hangi kurum, kuruluş ve kişilerle işbirliği yaptıkları belirtilmiştir.					
Kurum ve kuruluşların farklı ve benzer yönleri belirtilmiştir.					
Bu kuruluşların çevreye zarar verilmemesi için yaptıkları caydırıcı çalışmalarda ne ölçüde başarılı oldukları hakkında benzerlik ve farklılıklarına ait karşılaştırmalar yapılmıştır.					
DİL ve ANLATIM					
Rapor hâlindeki sunuda açık, anlaşılır ve akıcı bir dil ve anlatım kullanılmıştır.					
Toplam					
Toplam Puan					

Bu ölçekten alınacak en yüksek puan 40, en düşük puan ise 8'dir. Bu puan 100'lük not sistemine göre aşağıdaki formül kullanılarak nota çevrilebilir.

Örnek: Bir öğrenciniz ölçeğe göre 30 puan almış olsun.

$$\begin{array}{cc} 40 & 30 \\ 100 & x \end{array}$$

$$X = 75$$

75 puan 5'lik not sisteminde 70–84 aralığına gelir. Bu da 4'ün karşılığıdır.

Sample Rubric for the assessment of the Project. (Board of Education and Discipline, MoNE, 2011: 36) http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72