

Evaluation Process in English Language Teaching Program

 Dilek BÜYÜKAHISKA

Ondokuz Mayıs Üniversitesi, Eğitim Fakültesi, Yabancı Diller Bölümü

dbuyukahiska@gmail.com

Gönderilme Tarihi: 29/01/2019

Kabul Tarihi: 11/07/2019

Yayınlanma Tarihi: 30/07/2019

DOI: [10.30855/gjes.2019.05.02.003](https://doi.org/10.30855/gjes.2019.05.02.003)

Article Info

Anahtar Kelimeler:

Evaluation,
Assesment,
English Language
Teaching

ABSTRACT

In recent years, there has been a growing interest in the employment of effective evaluation and assessment techniques to meet the demands of 21st century foreign language instructional practices. A further shift has also taken place to initiate innovative and alternative assessment procedures in the continually changing and highly demanding foreign language teaching contexts. In a sense, this review study mainly focuses its discussion on evaluation process in English language teaching program. At first, some common terminology identified in the recent language evaluation literature are clarified elaborately. Next, an overview of the paradigm shift from traditional assessment to alternative assessment in English as a foreign language (EFL) context is elucidated. In particular, English language program followed in primary and secondary education in Turkey was another major issue to be held in research paper. Therefore, testing techniques for the assessment of language skills are discussed extensively in line with the requirements of English language program. Finally, necessary suggestions were advanced as to the ways to make English language teachers use the alternatives in language assessment.

İngilizce Öğretim Programında Değerlendirme Süreci

Makale Bilgileri	ÖZET
<p>Keywords: Değerlendirme, Ölçme, İngilizce dil öğretimi</p>	<p>Son yıllarda, 21. yüzyıl yabancı dil öğretim uygulamalarının taleplerini karşılamak için hangi etkili değerlendirme tekniklerinin kullanılması gerektiği konusuna ilgi artmaktadır. Sürekli değişen ve oldukça talepkar dil öğretimi ortamlarında çeşitli alternatif değerlendirme prosedürleri başlatmak için bir başka değişim daha da yaşanmıştır. Sürekli değişen ve oldukça talepkar yabancı dil öğretimi ortamlarında yenilikçi ve alternatif değerlendirme prosedürlerinin başlatılması yönünde bir başka değişim daha yaşanmaktadır. Bu anlamda, bu derleme çalışması ağırlıklı olarak İngilizce dil öğretim programındaki değerlendirme sürecine ilişkin tartışmaları ele almaktadır. Öncelikle, yabancı dil değerlendirme süreci alanyazınında tanımlanmış ve yaygın olarak kullanılan bazı terimler ayrıntılı bir şekilde açıklanmıştır. Daha sonra, İngilizce'nin yabancı dil olarak öğretildiği ortamlarda geleneksel değerlendirmeden alternatif değerlendirmeye geçiş paradigmasına genel bir bakış açıklanmaktadır. Özellikle, Türkiye'de ilk ve orta öğretimde izlenen İngilizce öğretim programı bu çalışmada ele alınan bir diğer önemli konudur. Bu nedenle, dil becerilerinin değerlendirilmesinde kullanılan teknikler İngilizce öğretim programının ihtiyaçları doğrultusunda tartışılmıştır. Son olarak, öğretmenlerin İngilizce öğretim programını değerlendirme sürecinde alternatif değerlendirme tekniklerini kullanmaları yönünde gerekli önerilerde bulunulmuştur.</p>

INTRODUCTION

In foreign language-teaching process, the terms test, assessment, evaluation, and measurement need to be clarified as they may cause ambiguity in literature. Although, these terms seem to have almost same meaning and to be used interchangeably, in fact, they are different in some way (Bachman & Palmer, 2010). Therefore, a number of researchers have attempted to highlight the intriguing concepts of evaluation process in language teaching.

It is slightly differentiated test from assessment. Assessment is a continuous process and may occur or performed at any time when the students respond a question, comment on something, share their opinions (Brown & Abeywickrama, 2010). To support that claim, Kunnan & Grabowski (2013) posit that assessment is performed at anytime and anywhere. On the other

hand, a test takes place at previously declared times and places. Learners, also, are aware that their responses are measured and evaluated (Brown, 2003). Testing is a kind of assessment and employed at the end of an instruction but assessment covers all of the educational system (Dickins, 2008).

The other terms confused are measurement and evaluation. Bachman (1990: 18) claims that “measurement is the process of quantifying the characteristics of persons according to explicit procedures and rules”. Namely, it is performed based on clearly defined rules and scheme. In a sense, there are quantifications such as numerical or letter grades and labels in the measurement. The quantifications provide teachers and institutes to compare students with each other. Additionally, mental traits and abilities including aptitude, intelligence, motivation, field independence/dependence, attitude, receptive skills are observed indirectly in measurement (Bachman, 1990). On the other hand, evaluation is a kind of systematic collection of outputs with the aim of making judgements by the interpretation of the measurement results. Evaluation is defined as “an attempt to understand what is going on to judge its worth and make decisions about it” (Desheng & Varghese, 2013, p. 33). Consequently, it is summarized the difference between these terms as: measurement and test contain quantification of monitoring. As a type of measurement, test is designed to draw out a specific sample of behaviour. The evaluation enables decision-making about the information provided by measurement (Bachman, 1990).

On the other hand, the difference between assessment and evaluation is that assessment is collection of information about learners’ language knowledge and learning but evaluation is interested in the overall programme not for individuals (Brindley, 2001). As White (1988, p. 149) states “evaluation focuses on issues, not on individuals”. In that sense, Richards (2001) classifies evaluation as formative, illuminative, and summative. The formative evaluation aims to find out how the program works and whether there are problems or not. Illuminative evaluation determines implementations to the programme in different aspects. Meanwhile, summative assessment seeks the efficiency of a programme. Similarly, Genesee (2001) states that the purposes of evaluation are to decide whether teaching and learning are suitable to the programme; to make decisions about learners’ status in a programme and to guide for teaching. In Figure 1, it is clearly presented the relationship between test, measurement, assessment, and evaluation (Brown and Abeywickram, 2010, p. 6)

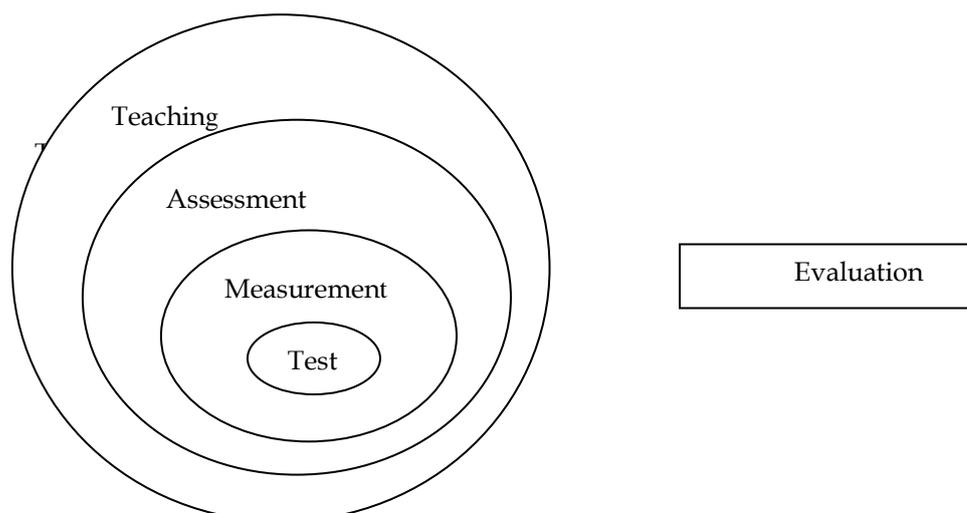


Figure 1. Test, measurement, assessment, teaching, and evaluation (Brown & Abeywickrama, 2010, p. 6)

As stated above, a test is a genre of assessment techniques. It is a method of measuring learners' knowledge or ability for a specific purpose. The characteristics of a test is listed by Brown & Abeywickrama (2010) as follows: a test is a method to measure performance in a given domain according to explicit rules but the results indicate competence. It is evident that assessment is a broad term including also testing. It provides information for guiding and informing instruction. The main purposes of assessment are selection, certification, motivation, diagnosis, accountability, and decision-making (Brindley, 2001). Testing is generally used for formal and standardized tests while assessment refers to more informal methods such as alternative assessment methods (Clapham, 2000). Kunnan & Grabowski (2013) note that the assessment contributes favourable, meaningful and reliable results, and promotes fair decisions.

There exist many types of assessment stated in related literature. Although, it is indicated that all assessment types are closely connected to each other in some way, there are no clear-cut distinctions among them. Table 1 fairly demonstrates the types of assessment. In the table, placing whether on the right or left column is not a sign of significance.

Table 1.

Types of Assessment (Council of Europe (CoE) (2001, p. 183)

Formal Assessment	Informal Assessment
Formative Assessment	Summative Assessment
Achievement Assessment	Proficiency Assessment
Norm- Referencing Assessment	Criterion- Referencing (CR) Assessment
Mastery CR Assessment	Continuum CR Assessment
Direct Assessment	Indirect Assessment
Performance Assessment	Knowledge Assessment
Subjective Assessment	Objective Assessment
Checklist Rating	Scale Rating
Impression	Guided Judgement
Holistic Assessment	Analytic Assessment
Series Assessment	Category Assessment
Assessment by Others	Self- Assessment

Informal assessment begins with unintended comments and praises (Brown, 2007). Therefore, this type of assessment is generally spontaneous. When teachers begin to make comments and respond to students' performances, it activates the assessment. That is to say, informal assessment is not data based but content and performance-based. It consists of the evidence, which teachers collect, in a classroom environment on a continuous basis to measure the progress of their students in the skills and content. It is also supported by Brown's statement (2007: 446) that informal assessment is embedded in classroom tasks designed to elicit performances, not for recording results and making judgements about students' competence. Different from informal assessment, formal assessment uses formal tests and structured continuous assessment. It obtains data promoted by results. This kind of assessment includes activities precisely designed to assess and record learners' performances and knowledge (Brown, 2003, 2007; Brown & Abeywickrama, 2010).

As for the distinction between formative and summative assessments, formative assessment is carried out to monitor students' progress and then necessary modifications are made according to results of the assessment (Hughes, 2003). Namely, formative assessment

directs attention on ongoing improvement (Richards, 2001), and it is conducted periodically for formative purposes to promote students' performance during the process of teaching (Katz, 2013). As Hughes (2003) suggests, students carry out a self-assessment to monitor their own improvement. On the other hand, summative assessment takes place at the end of the teaching process. It aims to measure what students achieve at the end of a course or unit (Brown, 2003). Formative assessment results are used to improve or revise the quality of instruction; however, summative assessment results give information about the teaching programme (Brindley, 2001). "It is used for summative purposes when they focus on what students have learned because of a period of instruction; it is assessment of learning." (Katz, 2013, p. 322).

Table 2.

Summative and Formative Assessment (Katz, 2013, p. 322)

Summative Assessment	Formative Assessment
Document Learning	Scaffold Learning
Diagnose Learning Needs	Provide Ongoing Feedback
Provide Information for Communication among Teachers, Students and Parents	Engage Students in Self - Assessment
Plan and Improve Instruction	

As for the difference between achievement assessment and proficiency assessment, achievement assessment is the assessing of the accomplishment of specific objectives. It is closely linked with the process of instruction (McNamara, 2000). As stated in Common European Framework of Reference (CEFR), it is course-oriented and exhibits and internal respective (CoE, 2001: 183). Specifically, teachers tend to use achievement assessment since it is closely related to the learners' participation. Based on its definitions by researchers (McNamara, 2000; Hughes, 2003; Brown, 2003; Brown & Abeywickrama, 2010), the features and purposes of achievement assessment are summarized as follows:

- Assessing level of competence
- Diagnosing strength and weaknesses
- Assigning grades
- Achieve certification
- Placement to higher educational institutes
- Curriculum evaluation

- Reliability
- Informational purposes

As another concept, proficiency assessment refers to the assessment of the abilities out of the classroom. Brown & Abeywickrama (2010) pose that proficiency assessment is not restricted to any courses and curriculum. Further, it merely deals with the real use of language ability (McNamara, 2000). It is oriented to real world ability and consists of both language knowledge and communicative competence (CoE, 2001).

The other type of assessment, norm-referencing assessment is a kind of assessment in which test takers are placed in rank order. It is about one's performance to that of others (Hughes, 2003). Each test taker's score is determined via statistic tools, such as mean, median, standard deviation, and percentile rank (Brown & Abeyickrama, 2010). In norm-referencing assessment, others' success is influential and scores are compared with each other. Conversely, criterion-referencing assessment indicates what test-takers know about language without telling something about others' performance (Hughes, 2003). It means that test-takers are assessed solely according to their abilities (CoE, 2001). Notably, Brown, & Abeywickrama (2010, s. 9) discuss "... criterion-referenced testing is of more prominent interest than norm-referenced testing". In addition, in the mastery criterion referencing assessment, a single minimum standard of competence or cut-off is settled to classify testees as masters or not-masters, with no degrees of quality in the achievement of objectives (CoE, 2001). Nevertheless, continuum criterion-referencing "... is an approach in which an individual ability is referenced to defined continuum of all relevant degrees of ability in the area in question" (CoE, 2001, p. 184). As a broad term, criterion referencing is labelled mostly as mastery learning or continuum whilst there are criteria that mean differently in that exam results in continuum criterion-referencing. These two types of assessments are used to determine language proficiency levels, such as A1, B2 or beginner, advanced etc.

As for direct assessment, it assesses what a candidate performs (CoE, 2001). According to Hughes' definition, direct assessment requires a candidate carry out a performance that is measured (2003, s. 17). Direct assessment is employed for productive skills (Hughes, 2003; Brown, 2003; Harmer, 2007; Brown & Abeywickrama, 2010). Hughes (2003, s. 17-18) asserts that direct assessment provides clear judgements about test-takers performance and creates positive washback. Indirect measurement, on the other hand, is used to measure a test taker's knowledge or ability that underlines the skill (Hughes, 2003). While direct measurement demands a real-life language use as much as possible, indirect measurement assesses a testee's knowledge about language (Harmer, 2007). Since direct measurement is limited to speaking and writing, indirect measurement offers a much wider range of assessment. The latter is superior to former one

(Hughes, 2003). Apart from these two kinds of assessment, there exist semi-direct assessment such as tape-recorded exams. The responses are recorded then scored (Hughes, 2003). Moreover, performance assessment is used to assess productive skills (writing, oral production, discussion, interactive tasks) instead of traditional paper-pencil tests (Brown & Abeywickrama, 2010). Performance means the production of language (CoE, 2001). The assessment indicates the behaviour that is intended to be measured (Brown, 2003). Abedi (2010) states that performance assessment demonstrates students' knowledge in the content area. It is used for diagnosis purposes to get data about test taker's knowledge. For a quality performance assessment, Salmani Nodushan (2008: 5) proposes the steps to be followed; deciding the skills and knowledge the students will acquire after the performance, designing performance tasks which students can show the skills and knowledge, creating performance criteria. In that sense, performance assessment includes authentic tasks, such as portfolios, projects, problem-solving tasks etc. (Brown, 2007). On the other hand, knowledge assessment requires answering questions so as to provide clues about test takers linguistic knowledge (CoE, 2001). Based on the assumptions, it is inferred that there is a similarity between performance assessment and direct assessment, and knowledge assessment and indirect assessment.

Another important distinction between assessment types is being objective or subjective. According to Hughes (2003, s. 22), "the distinction is between methods of scoring and nothing else". In other words, subjective assessment entails judgement on scoring, whilst there is no scoring in objective assessment. Furthermore, CEFR distinguishes "subjective assessment which is a judgement by an assessor" from "objective assessment which is an assessment in which subjectivity is removed" (CoE, 2001, s. 188). Subjective assessment may cause rater bias and unreliability of test; in contrast, objective assessment has fixed formats and increases reliability (Brown & Abeywickrama, 2010). When the subjectivity is reduced and objectivity is increased, the reliability is increased automatically. And in order to make a fair assessment, it should be as objective as possible (CoE, 2001).

Checklists are also a type of assessment tools used to determine whether the knowledge or ability exists. Rating scales are used, yet, to identify the quality or degree of the knowledge or skill. Checklists are simple assessment tools; there is no judgement on skills, while rating scales gives information about the quality and quantity of the skill (Erickson, 2011). Likewise, Richardson (2003) defines checklists as illustrating the completion of a task and rating scales as indicating rate of accomplishment. Also, CoE (2001, s. 189) defines the difference as in checklists, the emphasis is horizontal, namely how much of the content of the module successfully accomplished yet, in scales the emphasis is vertical, in other words how far up the scale goes.

Further, the impression is a completely subjective judgement on a performance, lacking any specific criteria (CoE, 2001). The main source of impression is tester's experience. In comparison, CEFR describes guided judgement as "... judgement with conscious assessment in relation to specific criteria" (2001, s. 189). It is concluded that since impression is entirely subjective, reliability is very low, so guided judgement increases reliability.

The discrimination between holistic and analytic assessment is similar to the distinction between impression and guided judgement. In a holistic assessment, the tester scores the overall process or product as a whole, without separating the component parts. In contrast, in an analytic assessment, the tester, first of all, scores separately individual parts of the product or performance and then sums the individual scores to obtain a total score (Mertler, 2001). In the analytic assessment, a tester assesses the components of a performance while in holistic one; tester assesses the performance or work as a whole. According to Hughes (2003) holistic assessment sometimes referred to as impressionistic assessment. Holistic assessment is a single scoring based on an overall impression and analytic assessment is scoring for each aspect of a task (Hughes, 2003). Brown & Abeywickrama (2010, s. 283) discuss the advantages of holistic assessment as fast evaluation, high inter-rater reliability. Similarly, Hughes (2003, s. 103) states the advantages of analytic assessment as it increases reliability, highlights sub-skills, and prevents ignoring any components.

As an analytic assessment, there is a single assessment task and a knowledge or ability is assessed in relation to the categories in an assessment grid in category assessment (CoE, 2001). However, in series assessment, there are certain separate tasks which are rated with a simple holistic grade scale.

In the learner-centered pedagogy, self- assessment is a kind of judgement about testees' his or her own proficiency. It is a feedback provided by test-takers about themselves, their abilities, what they can/cannot do. Self-assessment develops autonomy and intrinsic motivation. Brown & Abeywickrama (2010, s. 145-146) classify self-assessment into five categories:

- Direct assessment of performance
- Indirect assessment of competence
- Metacognitive assessment
- Socio-affective assessment
- Student generated tests

There are many advantages and disadvantages of this self-assessment. Gardner (2000, s. 51) shows an illustration to point its benefit for not only a learner but also teacher and institution.

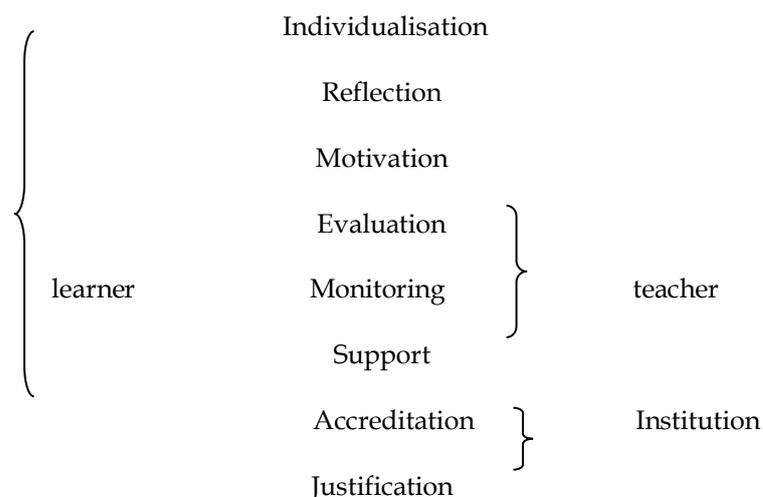


Figure 2. Benefits of self-assessment (Gardner, 2000: 51)

Additionally, unreliability, changing roles, lacking self-consciousness are the disadvantages of self-assessment (Gardner, 2000).

Apart from the classification and assessment types in Table 1, there are some other significant assessment types need to be discussed. They are called as large-scale testing vs. small-scale testing, continuous assessment vs. fixed-point assessment and discrete point testing vs. integrative testing.

Large scale language testing, known as a standardized test, is used for entrance to schools or graduating from there and for supervising test-takers progress with standardized development, administration, and scoring (Kunnan & Grabowski, 2013). The features of a large-scale testing are standard-based, a product of investigation, norm-referenced systematic scoring, and administration (Brown & Abeywickrama, 2010). The primary aspect of a large-scale testing is the uniformity (Kunnan & Grabowski, 2013). Table 3 shows the advantages and disadvantages of large-scale testing.

Table 3.

Advantages and Disadvantages of Large-Scale Testing (Brown & Abeywickrama, 2010: 106)

Advantages	Disadvantages
Easily available	Inappropriately use
Easily administered to large groups	Potential test biases
Streamlined scoring and reporting procedures	Not eliciting good sample of performance
Previously validated product	Mostly multiple-choice format

In contrast, Kunnan & Grabowski (2013) defines that a small-scale testing is used in an institution to monitor test-takers' progress, mostly, by a tester. The advantages of this testing are the items in the test which is directly relevant to the courses and the scoring procedure which can be arranged with the specific purposes and immediate feedback; still, less reliability is the main disadvantage of small- scale testing (Kunnan & Grabowski, 2013).

Furthermore, continuous assessment is a continuing assessment and assists teachers to evaluate a student (Iseni, 2011). It is based on observations about students' performance. A teacher carries out a continuous assessment with classroom performances, such as project, group works, and it is incorporated into the course (CoE, 2001). Thus, continuous assessment occurs during a course. Continuous assessment is ongoing process and requires collecting information regularly to find out what students know or learn. The students are observed periodically by teachers to be made a judgment about how well they do the tasks given (Du- Plessis, Prouty, Schubert, Habib, George, 2003).

Whereas, fixed-point assessment takes place on a particular day and students are evaluated according to the grades taken from an examination or another kind of assessment (CoE, 2001). In this assessment, the exam is important so the performance before the exam is not taken into consideration.

What is more, discrete point testing depends on structural linguistics theory (Demirezen, 2013). This kind of testing is developed on the presumption of language can be divided into its components and they can be assessed individually (Brown & Abeywickrama, 2010). The components are four skills and the units of language (phonology, morphology, lexicon, syntax, discourse). In addition, in this type of testing, each component is tested at a time and separately from the others (McNamara, 2000). In this type of testing, multiple-choice format is mostly used (McNamara, 2000; Demirezen, 2013, Brown & Abeywicrama, 2013). Different from discrete point testing, in an integrative test, skills are tested together (Demirezen, 2013). The reason of it is that

integrative tests are authentic and communication-based and there is a context within test items instead of isolated items (Brown & Abeywickrama, 2010). These tests measure the production (McNamara, 2000). Furthermore, cloze tests and dictation are mostly preferred test types in integrative tests (Brown & Abeywickrama, 2010; McNamara, 2000).

Alternative Assessment in ELT

Alternative assessment is accepted as a non-formal assessment method. In the alternative assessment, learners' products are assessed instead of what they remember or reproduce. As a reaction to standardized, paper-pencil test, alternative or authentic assessment was defined by McNamara (2001, s. 329) as a shift "away from the use of standardized multiple-choice tests in favor of more complex performance-based assessments". Alternative assessment is popular in recent years and its popularity is increasing day by day. Brown & Hudson (1998, s. 654-655) list the features of alternative assessment as:

- require students to perform, create, produce, or do something
- use real-world contexts or simulations
- are nonintrusive in that they extend the day-to-day classroom activities
- allow students to be assessed on what they normally do in class every day
- use tasks that represent meaningful instructional activities
- focus on processes as well as products
- tap into higher level thinking and problem-solving skills
- provide information about both the strengths and weaknesses of students
- are multi-culturally sensitive when properly administered
- ensure that people, not machines, do the scoring, using human judgment
- encourage open disclosure of standards and rating criteria
- call upon teachers to perform new instructional and assessment roles

Similarly, Hamayan (1995) states that alternative assessment presents authentic use of language, holistic and integrative view of language, learners' development process, and numerous addresses for learners' knowledge. In that sense, this type of assessment provides learners actual use of the target language as a means of communication since it reflects real-life communication contexts and situations outside the classroom (Beckman & Klinghammer; 2006).

Brown (2007) compares traditional assessment and alternative assessment in Table 4.

Table 4.

Traditional and Alternative Assessment (Brown, 2007, p. 462)

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

The most common and frequently used procedures for alternative assessment are portfolios, projects, journals, conferences, observations, interviews and simulations (Stoynoff, 2012).

As one of the most suggested alternative testing methods, portfolios are the collection of materials that are evidence of learning such as a wide range of samples including writing, notes, records, and projects (Dickins, 2008). A portfolio is an abridgement of a student's works indicating his or her learning (Davis & Ponnampereuma, 2005) and increases learner autonomy (Harmer, 2007). Since learners take their own responsibility in composing portfolios, they become autonomous (Hashemian & Fadaei, 2013). Apart from autonomy, portfolios increase learners' awareness of their learning (Yıldırım, 2013). Bryant & Timmins (2002, s. 10) list basic requirements of portfolio use:

- Teachers and administrators plan for and are trained in the portfolio approach to assessment
- Sufficient resources of time and energy must be allocated to support portfolio assessment
- Teachers must work as a team to plan for the implementation of portfolio assessment
- Parents and the public need to understand portfolio assessment
- The teacher's role is vital as a facilitator of the portfolio assessment.

- Portfolio assessment is a developmental process for both teachers and students
- Portfolio assessment provides a new perspective on learning for both teachers and students
- Self-evaluation of learning is an integral part of the portfolio process
- Collecting, selecting and reflecting on work is central to the portfolio process

According to Brown & Abeywickrama (2010), defining clear objectives, determining the type of works to include, deciding criteria to assess, periodic review and consulting, time, an appropriate place for portfolios and positive washback are the steps of portfolio assessment. However, Davis & Ponnampereuma (2005, s. 280) divide portfolio assessment into five stages: “collection of evidence of achievement of learning outcomes, reflection on learning, evaluation of evidence, defence of evidence and assessment decision”. The main advantages of portfolio assessment are to reinforce learning, motivation, and cooperation between teacher and learner, to sustain teachers’ position, to assess systematically besides authentic assessment, diagnosis learners’ weak and strong sides, providing autonomy, high validity, and beneficial washback; on the other hand, the main disadvantages are time-consuming, difficulty in determining and developing assessment criteria, finding sufficient and appropriate materials for a portfolio, low reliability (Brown & Hudson, 1998). Portfolios are suggested as the most frequently used testing techniques for the assessment of language skills (MEB, 2018).

Another alternative assessment method is rubric. A rubric is a scoring tool with criteria for a work and expresses the quality of the work on the base of these criteria from poor to excellent (Goodrich, 1997). Rubrics are easy to use, provide instructional feedback, and reinforce learning. They are, also, a clear indicator of the expectations about the work so that learners know the expectations and this increases the self- awareness (Andrade, 2000).

Projects, as a different method, are used to assess learners’ real-life language use. Projects provide a combination of using language knowledge and information on a specific topic. All kinds of study displaying learners’ knowledge about a topic are considered as a project and a project needs learner to employ a high level of thinking ability (Dikli, 2003).

In the same way, conferences are that the learners visit the instructor to discuss the works and learning the process. Conferences help learners to reflect their works and be aware of their learning. Both instructors and learners gather direct information about the learning process thanks to the conferences (Brown & Hudson, 1998). Richard and Renandya (2006, s. 351) state that conferencing is an effective form of oral teacher feedback. A short conference enables the teacher to ask the students about problematic parts of the letters writing. Kroll states the advantage of conferencing that it allows “the teacher to uncover potential misunderstanding that the student

might have about prior written feedback on issues in writing that have been discussed in class “(1991: 259).

Another alternative assesment approach practised in EFL classrooms to assess students’ English language skills is journal. Journals are daily or weekly writing entries by learners in which they reflect on their own learning experiences and progress (Abbas, 2012). As journal keeping is informal in nature, allows the student to get extensive writing practice (Richard & Renandya, 2002).

Table 5.

Testing Techniques for the Assessment of Language Skills (MEB, 2018)

Language Skills	Testing Techniques
Speaking	Collaborative or singular drama performances (Simulations, Role-plays, Side-coaching), Debates, Group or pair discussions, Describing a picture/video/story, etc., Discussing a picture/video/story, etc., Giving short responses in specific situations, Information gap, Opinion gap, Reporting an event/anecdote, etc., Short presentations, Talking about a visual/table/chart, etc
Listening	Different variations of matching (...the sentences with paragraphs ... pictures with the sentences, etc.), Discriminating between phonemes, Identifying interlocutors’ intentions and implicatures, Listen and perform/complete an action (E.g.: Listen and draw/paint, listen and match, listen and put the correct order, listen and spot the mistake, etc.), Listen and tick (the words, the themes, the situations or events, the people, etc.), Omitting the irrelevant information, Putting into order/reordering, Recognizing phonemic variations, Selective listening for morphological structure and affixation, True/False/ No information, Understanding overall meaning and supporting details, Recognizing specific information, Questions and answers.

Table 5.

Continue

Reading	Different variations of matching (...the sentences with paragraphs, ... pictures with the sentences, etc.), Finding specific information, Finding a title to a text, Identifying the gist and supporting details, Intensive reading, Read and perform / complete and action (E.g.: Read and guess the meaning of lexemes, Read and draw/paint, Read and solve the riddle), Solving a puzzle, Spotting text mechanics (reference, substitution, various types ellipses), True/False/No information, Transferring the text to a table/chart (Information transfer), Understanding the author's intention, Questions and answers.
Writing	Describing a picture/visual/video, etc., Filling in a form (hotel check in form, job application form, etc.), Note taking/ making, Preparing an outline, Preparing a list (shopping list, a to-do list, etc.), Reporting a table or a chart, Rephrasing, Rewriting, Writing short notes, entries and responses, Writing a paragraph/e-mail/journal entry/etc., Writing a topic sentence/ thesis statement
Alternative Assessment	Portfolio Assessment, Project Assessment, Performance Assessment, Creative Drama Tasks, Class Newspaper/Social Media Projects, Journal Performance, etc.

English Teaching Program

The English teaching program has been prepared based on the general goals of Turkish National Education as defined in The Basic Law on the National Education No. 1739, besides the Main Principles of Turkish National Education (MEB, 2015). In order to maintain effective foreign language education, course syllabi are revised periodically. Moreover, the transition from 8+4 model to 4+4+4 model leads to changes in English course program. According to the new model, English courses begin from the 2nd grade instead of 4th grade; in other words, students begin to learn English at younger ages. To put some evidence to this statement, Ersöz (2010) claims that exposing children to a foreign language accelerate their language acquisition process. They learn naturally, assimilating sound patterns, structures use, and usage of language similar to their first language. Necessary attempts were made to revise the foreign language program of all grades to maintain continuity (MEB, 2018). In that sense, the new English language-teaching program is based on the Common European Framework of References for Languages: Learning, Teaching, and Assessment (CEFR). CEFR emphasises real life language use in an authentic communicative

environment (CoE, 2001). In the new language teaching program, authentic use of language is supported, action-oriented approach is preferred to make students use the target language and improve their communicative competence (CoE, 2001; MEB, 2018). As language learning is a lifelong process, new program helps learners develop a positive attitude towards English. Apart from the authentic use of English, in this program there are drama and role-play activities, hand and craft activities, colouring activities for primary school students. In addition, they provide a positive and non- threatening learning environment (Elyıldırım & Ashon- Hayes, 2006).

In primary education, listening and speaking are emphasized; whereas reading and writing are incorporated in higher grades (MEB, 2018). At the 2nd, 3rd and 4th-grades, students develop their language skills through mostly listening activities, at the 5th and 6th grades, they began to read short texts, fill in the chart and write very short sentences in addition to listening and speaking activities. The 7th and 8th graders begin to read and write as an integral part of the language. In this program, there is no explicit grammar teaching. Namely, it is not structure based but communicative based (Bayyurt, 2013; Bayyurt & Alptekin, 2000).

As a result, it is maintained that learning a foreign language at an early age on a communicative base has advantages for children in terms of social and personal aspects. From social perspective, people using English fluently and accurately have a role of empowering nations' political and economic situations, improving a country's global power. Personally, children having an ability to communicate with foreigners enrich their lives, break the barriers and connect emotionally to the other people and their culture. (Ersöz, 2010). Foreign language teaching, therefore, plays a vitally important role in building peace around the World (European Centre for Modern Languages, 2000).

Table 6.
Model English Language Curriculum (MEB, 2018)

Levels [CEFR*] (Hours/Week)	Grades	Skill focus	Main activities/ strategies
1 [A1] (2)	2	Listening and Speaking	TPR/ Arts and crafts/Drama
		Listening and Speaking	
	3	Very Limited Reading and Writing	
		Listening and Speaking	
2 [A1] (4)	4	Very Limited Reading and Writing	Drama/Role play
		Listening and Speaking	
	5	Limited Reading	
		Very Limited Writing	
3 [A2] (4)	6	Limited Reading	Theme-based
		Very Limited Writing	
	7	Primary: Listening and Speaking	
		Secondary: Reading and Writing	
8	Primary: Listening and Speaking	Secondary: Reading and Writing	
	Secondary: Reading and Writing		

Table 6 summarizes English language curriculum followed in Turkey in terms of the levels, grades, skills and strategies to be used in English courses. Since this table is a common frame, EFL teachers may make necessary changes according to their students' needs, learning styles and strategies, abilities, proficiency levels, school type, and socio-economic situations. The table is prepared by the Ministry of National Education with regard to CEFR (MEB, 2018).

English Teaching Programs for Secondary School

After the transition from 8+3 model to 4+4+4 model, secondary school consists of 5th, 6th, 7th and 8th grades. As presented in Table 6, 5th and 6th grades are in A1 level while 7th and 8th grades are A2. The communicative use of language, cultural awareness, improving students' comprehension skill, self and pair assessment techniques are commonly used in all grades (MEB, 2018). It is inferred that this program is based on the spiral model of foreign language curriculum. Nation & Macalister (2010) state the benefits of this model as providing traced recycling of

subjects, lexical sets, vocabulary and language functions, helping learners remember their previous knowledge or learn better what they have not learnt sufficiently. The repetition of communicative functions in all grades provides long-term learning especially for foreign language learners.

In 1997, Turkish educational system experienced a number of fundamental changes. By these changes, compulsory education was increased from 5 years to 8. The Ministry of National Education introduced English from 4th grade. The main purpose of this radical change is to expose Turkish students to English at younger ages as much as possible for a longer period of time (MEB, 2001). Additionally, "Teaching English to Young Learners I -II " courses were implemented to the ELT program by Council of Higher Education (YOK). Hence, it is aimed that teachers of English should be well-qualified and experienced in teaching English to young learners. The YOK and MoNE heralded the significance of changing the pedagogical approach to making English courses more communicative and learner-centred (Kırkgöz, 2009).

English has been taught beginning from the 2nd grade since 2013 for better language learning to occur. The objectives and goals of English courses in primary schools are to increase students' consciousness of English, promote positive attitudes towards it, increase students' motivation, create stress-free and funny classroom environment to learn effective language learning, set up dialogues and meaningful contextualized learning (Kocaoluk & Kocaoluk, 2001) and help students be independent learners (İnceçay, 2012).

After the 4+4+4 system and teaching English beginning from the 2nd grades, teachers advocate the recent changes in the curriculum; however, some teachers need some in-service training related to the use of appropriate techniques for young learners. These teachers demand practical tips to apply theoretical language teaching principles to children. (Gürsoy, Korkmaz, & Damar, 2013). Also, it was suggested by Bayyurt (2013) that dividing English teachers as primary, secondary and high school teachers and training them according to the needs of the educational stages, increases qualification of the foreign language education in Turkey.

Assessment and evaluation are inextricably linked with language policy, language teaching methodology and curriculum design, teacher development. Although ELT program is claimed to design in communicative approach and aims to improve communicative competence, the assessment criteria may not meet the goals and objectives of English course, methods followed, materials selected and used in the curriculum. From this perspective, the mismatch causes either to ignore the program or the assessment. It is well known that the ignored one is the program since the tests force the teachers to teach to test (Popham, 2001). Based on related literature and teaching praxis, it can be stated that alternative assessment is assumed as a reaction to the traditional paper-pencil tests (Puppin, 2007). Shrestha (2013) asserts that alternative

assessment is to assess students' English language skills in a non-threatening class, and enables teachers to notice, appreciate and acknowledge the student's real potentials. Since, the traditional tests do not assess all language skills effectively. Moreover, the form of the question items, the test approach and techniques to be followed, test anxiety, socio-cultural and socio-economic differences among the students are the other crucial factors to be taken into consideration by EFL teachers in assessment of EFL achievement.

CONCLUSIONS

The suggested testing techniques for the assessment of English language skills should be employed in ELT program. The chosen techniques for each particular proficiency level, skills, and grade should be consistent with the curriculum. English language assessment and evaluation demand to follow diverse alternative and process oriented testing procedures (Portfolio Assessment, Project Assessment, Performance Assessment, Creative Drama Tasks, Class Newspaper/Social Media Projects, Journal Performance, etc.), self-assessment employed in order to encourage the students to monitor their own progress and achievement in the development of communicative competences (CoE, 2001; MEB, 2018). Formal evaluation is also be carried out through the use of written and oral exams, quizzes, homework assignments and projects by taking into consideration the proficiency level, grade or ages of the students. With respect to the technique, traditional techniques such as multiple-choice format gives detail information about neither receptive (reading, writing) nor productive (listening, speaking) skills. As a proper suggestion, technology may be highly implemented and integrated to test the knowledge and the competence in English language. Computer assisted language testing (CALT) is a beneficial method as it includes different sides of language testing and technology (Suvorov & Hegelheimer, 2013). Chapelle (2010) mentions three main motivating factors for using technology in language testing: efficiency, equivalence, and innovation. Technology provides effective testing and assessment of all language skills. The ongoing FATİH project and DynEd software are the examples of using technology in foreign language teaching. Similar programs are needed to provide authentic assessment. Solak and Avcı (2015) find out that DynEd program is sufficient for presenting information, exercises, testing, and evaluation. The students' English achievements and performances should be assessed and evaluated simultaneously with the help of interactive whiteboards and technology labs. Furthermore, as Woessmann (2002) states that individual creativity and freedom to choose the format may be provided to test English as a foreign language. Moreover, it is stated that foreign language assessment in Europe should follow the principles: communication, standardisation, and validation, certification, the relation to the CEFR, impact studies, support, and political issues (Eckes, Ellis, Kalnberzia, Pizorn, Springer, Szollas & Tsagari, 2005). It is clear that communicative approach is the main standard suggested

by EU and international foreign language tests. One point to note from the outset is that, in the modern information and communication technology (ICT) era, different skills such as global awareness, critical thinking, information and computer literacy are needed to implement alternative assessment in teaching rather than traditional ones. Together with the advancement in technology, EFL teachers adapt the changing assessment trends to their specific teaching context with authentic, communicative, multicultural and pedagogically appropriate materials. The new generation of web-based technologies, Web 2.0 tools should be integrated for the assessment purposes in ELT. Ching and Hsu (2011, p. 781) state the roles of Web 2.0 technologies as “automatic dissemination, powerful organization, enhanced interactivity and simplified collaboration”.

All in all, EFL teachers should be aware that language is a whole with its basic and sub-skills. Focusing on just one part of a language and neglecting the other parts prevents students from learning the language. EFL teachers should integrate all possible testing, assessment and evaluation processes based on the CEFR in their courses. In addition, all EFL teachers examine English teaching program, study on it in details, and design the courses according to both the curriculum and the needs of their students. Effective assessment and testing techniques that comply with the ELT program should be exploited by EFL teachers as much as possible. Practically, together with English teaching program, teachers should follow the new trends in foreign language teaching assessment and testing process in order to adopt the 21st century instructional goals with the digital native students’ needs. The innovations bring new paradigm and plentiful enjoyable activities to the courses. Further, technology provides plentiful sources for English teaching. Teachers should use technology whenever possible. Therefore, foreign language teacher training programs should be monitored, revised and regularly modified to cope with the challenges of language assessment in specific language contexts. In order to be able to employ the most effective assessment or evaluation techniques, EFL teachers’ views, needs, expectations, and suggestions should be analyzed comprehensively. Teachers need to be well-experienced and skillful enough when and how to use which tool for effective and efficient language teaching and learning.

REFERENCES

- Abbas, Z. (2012). Difficulties in using methods of alternative assessment in teaching from Iraqi instructors’ points of view. *AL-Fatih Journal*, 48, 23-45.
- Abedi, J. (2010). *Performance assessments for English language learners*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.
- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.

- Bachman, L. & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. New York: Oxford University Press.
- Backman, L. O. and Klinghammer, S. J. (2006). *Shaping the way we teach English instructor's manual*. Washington: University of Oregon.
- Bayyurt, Y. & Alptekin, C. (2000). EFL syllabus design for Turkish young learners in bilingual school contexts. J. Moon & M. Nikolov (Eds.). *Research into teaching English to young learners* (pp. 312-322). Pécs: Pécs University Press.
- Bayyurt, Y. (2013). Foreign language learning for young learners in 4+4+4 education system. Sarıçoban, A. & Öz, H. (Eds.), *Hacettepe Üniversitesi Eğitim Fakültesi İngiliz Dili Eğitimi Anabilim Dalı 1. Yabancı Dil Eğitimi Çalıştayı Bildirileri* (Proceedings of the 1st National Workshop on Foreign Language Education) (pp. 115-125). Ankara: Hacettepe University Publishing House.
- Brindley, G. (2001). Assessment. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to speakers of other languages* (pp.137-144). Cambridge: Cambridge University Press.
- Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy*. (3rd Ed.). New York: Pearson Education, Inc.
- Brown, H.D. & Abeywickrama, P (2010). *Language assessment: Principles and classroom practices*. (2nd Ed.). New York: Pearson Education, Inc.
- Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.
- Brown, S. (2003). Assessment that works at work. *The Newsletter for the Institute of Learning and Teaching in Higher Education*. 11, 6-7.
- Bryant, S. L., & Timmins, A.A. (2002). *Portfolio assessment: Instructional guide* (2nd Ed.). Hong Kong: Hong Kong Institute of Education.
- Chapelle, C. A. (2010). Technology in language testing [video]. Retrieved from <http://languagetesting.info/video/main.html>
- Ching, Y. H., & Hsu, Y.-C. (2011). Design-grounded assessment: A framework and a case study of Web 2.0 practices in higher education. *Australasian Journal of Educational Technology*, 27(5), 781-797.
- Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics*, 20, 147-161.
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Cambridge University Press.

- Davis, M. H. and Ponnampereuma, G. G. (2005). Portfolio assessment. *Journal of Veterinary Medical Education*, 32(3), 279-284.
- Demirezen, M. (2013). Testing. In A Sariçoban (Ed.), *Öğretmenlik alan bilgisi testi: İngilizce öğretmenliği* (p. 165- 194). Ankara: Murat Yayınları.
- Desheng, C., & Varghese, A. (2013). Testing and evaluation of language skills. *IOSR Journal of Research & Method in Education (IOSR-JRME)*, 1(2), 31-33.
- Dickins, P. R (2008). Classroom assessment. In T. Hedge (Ed.), *Teaching and learning in the language classroom*. (pp. 375-403). Oxford: Oxford University Press.
- Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *The Turkish Online Journal of Educational Technology - TOJET*, 2(3), 13-19.
- Du-Plessis, J., Prouty, D., Schubert, J., Habib, M., & George, E. S. (2003). *Continuous assessment. A practical guide for teachers*. Washington, DC: American Institute for Research.
- Eckes, T., Ellis, M., Kalnberzia, V., Pizorn, K., Springer, C., Szollas, K., & Tsagari, C. (2005). Progress and problems in reforming public language examinations in Europe: Cameos from the Baltic States, Greece, Hungary, Poland, Slovenia, France and Germany. *Language Testing*, 22(3), 355-377.
- Elyıldırım, S., & Ashton-Hay, S. (2006). Creating positive attitudes towards English as a foreign language. *English Teaching Forum*, 44(4). Retrieved from <http://eprints.qut.edu.au/12256/>
- Ersöz, A. (2010). Teaching English to young learners. Ankara: EDM. *European centre for modern languages of the council of Europe*. Retrieved from <http://www.ecml.at/>
- Gardner, D. (2000). Self-assessment for autonomous language learners. *Links & Letters*, 7, 49-60.
- Genesee, F. (2001), Evaluation. In Carter, R., Nunan, D. (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 144-150). Cambridge: Cambridge University Press.
- Goodrich, H. (1997). Understanding rubrics. *Educational Leadership*, 54(4), 14-18.
- Gürsoy, E., Korkmaz, Ş. Ç., & Damar, E. A. (2013). Foreign language teaching within 4+4+4 education system in Turkey: Language teachers' voice. *Eurasian Journal of Educational Research*, 53, 59-74.
- Hamayan, E.V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics*, 15, 212-226.
- Harmer, J. (2007). *The practice of English language teaching* (4th Ed.). Pearson Education Limited.
- Hashemian, M. & Fadaei, B. (2013). Fostering EFL learners' autonomy in light of portfolio assessment: Exploring the potential impact of gender. *Iranian Journal of Language Teaching Research*, 1(2), 135-151.

- Heaton, J. B. (1990). *Writing English language tests*. New York: Longman Group UK Limited.
- Hughes, A. (2003). *Testing for language teachers* (2nd Ed.). Cambridge: Cambridge University Press.
- İnceçay, G. (2012). Turkey's foreign language policy at primary level: Challenges in practice. *International Association of Research in Foreign Language Education and Applied Linguistics ELT Research Journal*, 1(1), 53-62.
- Iseni, A. (2011). Assessment, testing and correcting students' errors and mistakes. *Language Testing in Asia*, 1(3), 60-90.
- Katz, A. (2013). Assessment in second language classrooms. In M. Celce- Murcia, D. M. Brinton & M. A. Snow (Eds.), *Teaching English as a second or foreign language*, (4th Ed.) (p. 320-340). Boston, MA: Heinle ELT.
- Kırkgöz, Y. (2009). Globalization and English language policy in Turkey. *Educational Policy*, 23(5), 663-684.
- Kocaoluk, F., & Kocaoluk, M.S. (2001). *İlköğretim okulu programı 1999-2000 [Primary education curriculum 1999-2000]*. İstanbul: Kocaoluk Publishers.
- Kroll, B. (1991). *Teaching writing English in the ESL context in Murica, teaching English as a second language*. Newyork: Newburry house
- Kunnan, A. J., & Grabowski, K. (2013). Large- scale language assessment. In M. Celce-Murcia, D. M. Brinton & M. A. Snow (Eds.), *Teaching English as a second or foreign language*. (4th Ed.) (p. 204-320). Boston, MA: Heinle ELT.
- McNamara, T. (2000). *Language testing*. New York: Oxford University Press.
- MEB (Milli Eğitim Bakanlığı), (2013). Orta öğretim kurumlarına geçiş yönergesi, birinci kısım, tanımlar, madde: 4. *Tebliğler Dergisi*, Kasım 2007/2602.
- MEB (Milli Eğitim Bakanlığı). (2018). *İlköğretim kurumları (ilkokullar ve ortaokullar) İngilizce dersi öğretim programı*. Ankara. <http://ttkb.meb.gov.tr/program2.aspx?islem=1&kno=214>
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25).
- Nation, I.S.P. & Macalister, J. (2010). *Language curriculum design*. New York: Routledge.
- Popham, J. W. (2001). Teaching to the test. *Educational Leadership*, 58(6), 16-20.
- Puppın, L. (2007). A paradigm shift: From paper-and-pencil tests to performance-based assessment. *English Teaching Forum*, 45(4),10-17.
- Richard, J. C., & Renandy, W. A. (2002). *Methodology in language teaching*. Cambridge. Cambridge University Press.

- Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge: Cambridge University Press.
- Richardson, E. (2003). *Scoring performance assessments: Checklists, rating scales and rubrics*. <http://www.alabamapepe.com/profdevmodule/scoring/scoring.pdf>
- Salmani Nodoushan, M. A. (2008). Performance assessment in language testing. *Journal on School Educational Technology*, 3(4), 1-7.
- Shrestha, P. (2013). Alternative assessment approaches in primary English language classrooms. *Journal of NELTA*, 18(1-2) 148-163.
- Solak, E. & Avcı, D. (2015). Evaluation of Dyned software language program from teachers' perspective. *International Journal of Social Science*, 31, 235-247.
- Stoyhoff, S. (2012). Looking backward and forward at classroom-based language assessment. *ELT Journal*, 66(4), 523-532.
- Suvorov, R., & Hegelheimer, V. (2013). Computer-assisted language testing. In A. J. Kunnan (Ed.), *Companion to language assessment* (p. 593-613). Malden, MA: Wiley Blackwell.
- White, R. V. (1988). *The ELT curriculum. design, innovation and management*. Oxford: Basil Blackwell Ltd.
- Woessmann, L. (2002). How central exams affect educational achievement: International evidence from TIMSS and TIMSS-repeat. *Harvard University program on education policy and governance working*, Paper No. PEPG/02-10.
- Yıldırım, R. (2013). The portfolio effect: Enhancing Turkish ELT student-teachers' autonomy. *Australian Journal of Teacher Education*, 38(8), 92-110.