Commonly Used Techniques in Testing Foreign Language Skills and Communicative Language Teaching

Koray KAÇAR[1]

ABSTRACT

Teaching and testing foreign language skills are long and challenging processes that cannot be separated from each other. Many traditional test items and alternative methods are utilized for various testing purposes. However, for a successful evaluation, alternative method and traditional assessment types should be handled in communicative language teaching for a successful evaluation. The basics of Communicative Language Teaching are important to increase foreign language learners' communicative competence. In this sense, this review illustrates traditional testing item types after defining the difference between test and assessment. Secondly, testing language skills and components are put under a critical perspective. Then, the basics of Communicative Language Teaching are explained. Eventually, suggestions about adopting all testing types and methods for an effective assessment are discussed. Also, technology-based assessment tools are recommended for an authentic assessment. Correspondingly, the review mainly tries to shed light on testing language skills and language components. Moreover, it aims to help foreign language teachers prepare exams that follow the basic principles of Communicative Language Teaching.

Anahtar kelimeler: Testing, Assessment, Test techniques, Communicative language teaching, Testing language skills

Yabancı Dil Becerilerinin Ölçülmesinde Yaygın Olarak Kullanılan Teknikler ve İletişimsel Dil Öğretimi

ÖZET

Öğretme ve ölçme, yabancı dil becerilerinin öğretiminde birbirinden ayrılamayan uzun ve zorlayıcı süreçlerdir. Farklı ölçme amaçları için kullanılan pek çok farklı ölçme yöntemi bulunmaktadır. Ancak İletişimsel Dil Öğretiminde başarılı bir değerlendirme yapılabilmesi açısından geleneksel ölçme yöntemlerinin yanı sıra alternatif değerlendirme yöntemleri de kullanılmalıdır. İletişimsel Dil Öğretiminin

[1] Yüksek Lisans Öğrencisi, Ondokuz Mayıs Üniversitesi Lisansüstü Eğitim Enstitüsü, Samsun, Türkiye, koray_kacar22@hotmail.com. ORCID ID: 0000-0002-0258-0016

temel ilkeleri yabancı dil öğrenenlerin iletişimsel yeterliklerinin artması için önemlidir. Bu derleme çalışmasında öncelikli olarak ölçme ve değerlendirmenin farkı tanımlanmıştır. İkinci olarak, sırasıyla dil becerilerinin ve öğelerinin ölçülmesi ele alınmıştır. Daha sonra İletişimsel Dil Öğretimi ilkeleri açıklanmıştır. Son olarak da etkili bir değerlendirme için tüm ölçme yöntem ve tekniklerinin kullanılması önerilmiştir. Ayrıca otantik değerlendirme için teknoloji temelli ölçme araçlarının kullanılması da önerilmiştir. Tüm bunlar bağlamında, bu derlemenin amacı dil becerileri ve dilin öğelerinin ölçülmesi konusunda açıklama yapmaktır. Ayrıca yabancı dil öğretmenlerinin İletişimsel dil Öğretiminin temel ilkelerine uygun sınavlar hazırlamaları için yardım etmek de hedeflenmiştir.

Keywords: Ölçme, Değerlendirme, Ölçme teknikleri, İletişimsel dil ğretimi, Dil becerilerinin ölçülmesi

1. INTRODUCTION

Testing English proficiency plays an important role evaluating the outputs of investments and programs in teaching English. In this regard, testing is a kind of mirror. In other words, it demonstrates the quality of education provided in English classes because testing is a crucial field on the grounds that teaching and testing are linked to each other. In this regard, studying testing without referring to teaching is impossible, or vice versa (Heaton, 1990). Testing English proficiency is not considered an isolated process. That is to say, language testing cooperates with the studies in language teaching (Bachman, 1990).

1.1. Test and Assessment

In EFL teaching, the terms test and assessment cause confusion and need explication. Although these terms are often considered the same and used interchangeably, they are different (Bachman & Palmer, 2010). Brown and Abeywickrama (2010) differentiate assessment from tests as assessment is a continuous process. Learners' opinions and problem-solving abilities can be assessed anywhere and anytime (Kunnan & Grabowski, 2013). However, the test is a way to measure and assess learners' abilities and knowledge (Brown, 2007). Moreover, a test occurs at specific and determined times and places, and testees know that their answers are measured and evaluated (Brown, 2003). A test is a fragment of assessment to measure learners' knowledge and abilities (Brown & Abeywickrama, 2010). In this sense, assessment is broader than testing.

1.2. Types of Test Items

As one of the most preferred types of questions, *multiple choice* requires testees to choose the correct or the most suitable answer for the question or statement. Multiple-choice items consist of a question and the alternatives or the choices as a response and test takers are asked to choose one of the choices that suit the question or problem (Clay, 2001). They are considered one of the most objective types and are mostly used in testing (Cunningham, 1998). These test items recall information (Cunningham, 1998; Clay, 2001). The main advantages of the multiple-choice test items are being practical and reliable (Hughes, 2003). The scores obtained through this testing type are also considered reliable, rapid and economical (Demirezen, 2013). On the other hand, Hughes (2003) points to the difficulties of this technique. In this type, mostly recognition knowledge is tested, in other words, testers may guess the correct answer or find it out by chance. Moreover, writing a valid and reliable test item is difficult, and negative washback may occur. Similarly, Cunningham (1998) lists four important learning outcomes that cannot be tested with the multiple-choice format: being original, organizing information, transferring the information into action, and presenting information. Multiple-choice question items are used because they are

the most objective, practical and reliable. They can test learners' academic knowledge. However, their preferable advantages, communicative competence and productive skills cannot be assessed with multiple choice question items.   An example of a multiple-choice item looks like (adapted from British Council):

Cook is _____today for being one of Britain's most famous explorers.
   a) recommended   b) reminded   c) recognised   d) remembered

In *true / false questions*, there is a statement and test-takers decide whether the statement is true or not. This kind of question is good for knowledge level (Clay, 2001). The main advantage of true/false items is that they can be completed quickly and scored easily (Cunningham, 1998). However, the main drawback is that test-takers have a 50 % chance of choosing the correct option (Hughes, 2003). Some modifications can be made to minimize the chance factor. One of these changes is adding a third option such as 'it is not stated' (Alderson et al., 1995). The other one is to ask for an explanation of their choice (Hughes, 2003). An example of this type is like this (adapted from British Council):

England won the World Cup in 1966.   T/F

In *cloze tests*, testees are given sentences with gaps and testees are expected to fill in the blanks. Cloze tests combine language skills and general linguistic ability (Heaton, 1990). According to Alderson (1996), the cloze test is classified into two groups: (i) pseudo- random and (ii) rational cloze procedure. The rational cloze procedure is also called gap-filling (Cohen, 2001; Ozerova, 2004). In gap- filling tests, words are omitted in view of predetermined linguistic criteria and macro-level discourse links (Cohen, 2001). However, in a pseudo-random procedure, items are deleted in a systematic way (Ozerova, 2004). According to Clay (2001), the main advantages of using cloze tests are that they facilitate comprehension, minimize the chance factor, and increase intensive study. On the other hand, the disadvantages are that cloze tests overemphasize memorization, and gaps are mostly limited to one or two words. Cloze tests increase practicality, validity, and reliability; but are limited to assessing learners' knowledge levels. An example is for the cloze test is (adapted from British Council):

Complete the text by adding a word to each gap.

This is the kind ____ test where a word ____ omitted from a passage every so often. The candidate must ____ the gaps, usually the first two lines are without gaps.

In *matching* item tests, two related items are matched. Matching items are comprised of two parts, one of which is stem (question), and the other is responses (Clay, 2001). The matching items are used mostly to assess testees' skill for matching terms and definitions (Cohen, 2001), "phrases with other phrases", and "problems with solutions" (Clay, 2001, p.27). Apart from words and terms, sentences and visual materials can be used as matching items (Demirezen, 2013). The example for matching items is like the following (adapted from British Council):

Match the word on the left to the word with the opposite meaning.

fat	old
young	tall
dangerous	thin
short	safe

Paraphrase is expressing something in a different way (Callison-Burch, 2007). In this testing type, testees are asked to re-write the sentences without any changes in the meaning (Hughes, 2003; Harmer, 2007). This type is useful to get information about testees' knowledge of language

systems (Harmer, 2007). In addition, it is used to assess one's oral ability as "orally receiving and orally relaying a message" (Brown and Abeywickrama, 2010, p.204). Paraphrase question items show whether learners reach the comprehension level. Comprehension involves understanding facts and ideas by organizing, summarizing, translating, generalizing, and giving descriptions without changing the main ideas. Testees' are asked to re-state the given statements with their sentences, so they need to comprehend the statements. As an example (Harmer, 2007, p. 383):
I am sorry that I did not get her an anniversary present.
I wish ……………………………………………….........

Short-answer item tests require fulfilling the tests with short and limited responses. Short answer items are short sentences half of which are missing or a statement to be filled (Cunningham, 1998; Demirezen, 2013). Short answer items are very useful for checking testees' vocabulary knowledge. They show what testees know exactly but they are subjective and suitable for just knowledge level (Cunningham, 1998), and not appropriated for complex learning outcomes (Saeed and Noor, 2011). Scoring keys should be prepared in advance in order to provide objectivity and increase reliability (Valette, 1977).

In an *essay* item test, test-takers are requested to write about a specific topic or a particular situation. In such exams, test-takers are expected to produce longer responses than any other test items, and the length may vary from sentence to page (Cunningham, 1998). For answering essay format items, test-takers need to use skills such as critical thinking (Demirezen, 2013); thus, the evaluation should be made carefully (Brown and Abeywickrama, 2010). Cunningham (1998) states creativity, organization ideas, and different ways of expressing ideas as the advantages of this type of question while the most obvious disadvantage is low-reliability (Cunningham, 1998, Demirezen, 2013).

1.3. Testing Language Skills

Listening comprehension is understanding the sounds in a context (Akıncı, 2010). Moreover, listening is a prerequisite for oral communication and for this reason Brown (2003), Hughes (2003), and Brown and Abeywickrama (2010) recognize listening as an essential part of verbal communication. Listening is a receptive skill; thus, indirect testing methods are used to assess it. According to Brown (2007), clustering, redundancy, reduced forms, performance variables, colloquial language, the rate of delivery, interaction, stress, rhythm, and intonation make listening tests difficult.

Buck (2001) proposes that the most suitable way of producing a construct of listening is by combining competence-based model and task-based model. This combination provides the ability "to process extended samples of realistic spoken language, to understand the linguistic information and to make whatever inferences are unambiguously implicated by the content of the passage" (p. 114). In addition, Buck (1998) categorizes listening tests into two groups: process approach and product approach. The orocess approach tries to identify the sub-skills in listening and then assesses the testees whether they master them. On the other hand, in the product approach, testees are given a passage to listen to and then they are assessed through this passage.

Weir (1993) divides listening comprehension test requirements into four categories: listening for direct meaning, inferred meaning, contributory meaning and listening for note taking. In the first requirement, gist, details, and manner of the speaker are examined. In the second one, making inferences and inquiries, recognizing the communicative and social functions of utterances are

scanned. In the third one, linguistic features are checked. In the last requirement, summarizing and paraphrasing the text and selecting relevant key points are underlined.

Brown (2007, p.309-310) lists the listening comprehension methods as:
• Intensive listening tasks: distinguishing phonemic, morphological pairs, repetition, stress patterns, and recognizing paraphrases
• Responsive listening tasks: immediate responses to questions
• Selective listening tasks: cloze procedures and information transfers
• Extensive listening: dictation, lectures, dialogues, narrations

Hughes (2003) suggests multiple choice, short answer, gap filling, note-taking, partial dictation, and transcription as listening test techniques. Moreover, Richards (2006) proposes two kinds of listening activities. They are noticing activities and restructuring activities. Noticing activities serve for comprehension activities while restructuring activities for productive use of listening texts. On the other hand, Madsen (1983) divides the listening test items into three classes: limited response, multiple choice appropriate response, and extended communication.

Speaking is a productive skill and testees' performance can be observed directly. The aim of testing speaking is to observe the ability to communicate in the target language (Hughes, 2003). As Brown and Abeywickrama (2010) state there is a close relationship between speaking and listening, so listening affects testees' oral production. Thus, speaking ability includes both comprehension and production (Hughes, 2003).

Madsen (1983) mentions some difficulties in testing speaking skills, such as testing fluency or accuracy, motivating students to speak, testing each student individually, and evaluating many things simultaneously. O'Sullivan (2008) suggests that the writer of a speaking test should be careful about the construction of definition, the predictability of responses, the interlocutor effect, validity, and reliability. On the other hand, Brown (2007) clarifies the advantages of speaking tests as observable, recordable, and measurable. As Gonzalez (1996) states, despite the difficulties in scoring and administration, testing speaking is necessary and important, especially in the communicative use of language.

Lazaraton (2001) distinguishes oral skill assessment into two categories. The first one is classroom performance and the second one is a large-scale oral examination. Classroom performance includes various classroom activities, such as extended chunks and visual or interactive chunks. Different from classroom performance, large-scale oral examination is a requirement to get into some universities or some institutions and fluency, accuracy, coherence, pronunciation are important in this type of examination (Lazaraton, 2013).In terms of fluency and accuracy, testing fluency attaches importance on speaking without too much hesitating and stopping while testing accuracy emphasises on correct grammar and vocabulary (Riddell, 2003).

Similar to the types of listening, Brown (2007) expresses the types of speaking as imitative, intensive, responsive, interactive, and extensive. Imitative tasks require just imitations and repeats. There is no need to understand and interact. In imitative speaking, pronunciation and prosody are important (Brown and Abeywickrama, 2010). In intensive tasks, reading aloud, directed response, oral cloze procedure, and translation are commonly used methods (Brown, 2007). Responsive tasks include interaction and comprehension. There are conversations, role-plays, games and small talks (Brown and Abeywickrama, 2010). As different from responsive

tasks, interactive tasks need longer and more complex interactions between participants. Expensive tasks, on the other hand, include lectures, presentations, and long speeches (Brown, 2007).

Madsen (1983) proposes holistic scoring and objectified scoring for speaking skills. Holistic scoring is used when many criteria are evaluated at the same time, whereas objectified scoring is made with a checklist and, each aspect of the task is scored independently. Hughes (2003) states holistic scoring is the assessment base on the overall impression. The main advantage of this scoring is that it saves time because it is performed fast. The other scoring type is analytic. It separates the tasks and checks separately, which makes scoring reliable but it is time- consuming. Testing speaking skills heavily depends on the authentic use of language. One of the main reasons to learn a foreign language is communication. As a result, both the use and usage of language are important for assessing listening skills.

Reading is considered the most important skill and people have the opportunity to read English materials, but they do not have the chance to communicate orally (McDonough et al., 2013). Similar to listening, reading is a receptive skill and the process or product of reading cannot be observed directly (Brown and Abeywickrama, 2010). However, reading is more complex than listening in that testees must develop bottom-up, and top-down strategies, and suitable schemata as an efficient reading comprehension (Brown and Abeywickrama, 2010).

As Brown (2007) states, for a good reading test, micro or macro skills, the genre of reading, and the type of reading should be determined. In the taxonomy of reading types, each reading type serves different purposes. For example, a perceptive reading task involves components of discourse such as letters, words, and punctuation. However, selective reading is the realizing basic grammatical, lexical or discourse features, or language with short texts and visuals. On the other hand, an interactive task focuses on relevant features between paragraphs and organizing them in a logical order. Finally, the extensive task includes articles, essays, reports, and books (Brown and Abeywickrama, 2010). According to Brown (2007) the tasks used in receptive reading are reading aloud, copying, multiple-choice, and picture-cued identification. Cloze, matching, gap-filling, and editing tasks are suggested for selective reading. Different from the other tasks, interactive and extensive tasks such as skimming, scanning, re-ordering, summarizing, outlining, responding, etc., require complete understanding of the text.

Moreover, Kitao and Kitao (1996a) divide reading comprehension questions according to the level of the testees: testing low-level skills, testing middle or higher levels. Testing low-level skills include word recognition, sentence recognition, and matching the word with a picture. In the testing middle or higher levels, the techniques are true-false questions, multiple-choice questions, short answer completion, and ordering. Hughes (2003) defines the criteria that a reading text suits for the assessment of reading skills as type, form, graphic features, topic, style, readability, length, lexis, and structure. On the other hand, Kitao and Kitao (1996a) identify the difficulty, variety, type of text and background knowledge for choosing a text in assessment. Hughes (2003) warns that in scoring reading ability, including testees' grammar or vocabulary mistakes in assessment decreases validity.

Both teaching and testing *writing* are complex processes, as writing ability requires proficiency in grammatical, oratorical, lexical, and conceptual items (Heaton, 1990). Similarly, Kitao and Kitao (1996b) describe the elements of writing as grammatical ability, lexical ability, mechanical ability, stylistic skill, organizational skill and judgements of appropriateness. Writing is a productive skill

and it is tested directly (Hughes, 2003). In process-oriented instruction, testing writing is a prickly matter (Brown, 2007). The assessment of writing can be formative and summative. The assessment is formative when the process-oriented approach is used while it is summative if there is a final product (Brown, 2007).

Brown and Abeywickrama (2010) divide writing tasks into four categories: Imitation, intensive, responsive, and extensive writing. Imitative writing is producing letters, words, and short sentences. Intensive writing is creating a suitable vocabulary in a context and linguistic elements in the sentences. The form is more important than meaning in intensive writing. However, meaning is important in responsive writing, and testees are expected to produce meaningful and logical sentences or coherently ordered paragraphs. Eventually, extensive writing means using all writing strategies and writing for all purposes and types. Brown (2007) defines the tasks for each writing type. The tasks for imitative writing are handwriting letters, copying, spelling, simple word writing. Similar to imitative writing tasks, intensive writing task is form focused and includes dictation, dicto-comp, ordering, and transformation. Different from these activities responsive tasks are meaning focused, such as paraphrasing, construction, and responding. Moreover, extensive tasks are writing in all types and genres.

Hughes (2003) suggests two kings of scoring: holistic and analytic scoring. As stated in testing speaking skills, holistic or impressionistic is based on the general impression of the writing product. It is fast and includes inter-rater reliability (Brown and Abeywickrama, 2010). On the other hand, in analytic scoring, the elements composing the written product are graded separately. Apart from these scoring types, Brown and Abeywickrama (2010) define another scoring type called primary trait scoring. Primary trait scoring focuses on the purpose of the writing and whether it fulfils the purpose (Brown and Abeywickrama, 2010). For instance, if the purpose of the paragraph or essay is to dissuade from something, the score is given as much as the testees fulfil this purpose. Holistic and primary trait scoring provides little or no washback but analytic scoring provides positive washback since it highlights the testees' both strong and weak sides.

Furthermore, correction of writing drafts or feedback is not just teachers' responsibility. There are three kinds of correction such as teacher correction, peer correction, and self-correction (Riddell, 2003). Teacher correction provides fast, efficient and accurate correction while discouraging independence and demotivating testees (Riddell, 2003). Teacher correction on writing can be written, in individual meetings or recorded spoken feedback (Weigle, 2013). Another correction is peer correction, which enables students to be involved in the correction process and provides authentic feedback (Weigle, 2013). Peer correction increases critical reading skills, and peers can find something that the teacher misses. Moreover, it creates cooperation between students (Riddell, 2003; McDonough et al., 2013). Finally, self-correction increases students' independence since they are actively involved in the correction process. The correction of written works can be an immediate or delayed correction (Riddell, 2003). Also, in the correction symbols (Weigle, 2013), class correction and grading can be used (Riddell, 2003).

Language skills cannot be tested completely independently from each other; there is a close relationship between all language skills. Thus, foreign language teachers should employ different methods and techniques simultaneously for an accurate assessment.

1.4. Testing Language Components

In the traditional language teaching approaches, *grammar* is regarded as the core of the language. Brown and Abeywickrama (2010, p. 293) state "… knowing a language meant knowing the grammatical structures of that language." In the traditional approach, grammar knowledge is assessed by decontextualized, discrete-point items such as re-ordering, cloze procedure, error correction, completion, and sentence combination (Larsen-Freeman, 2009). Thanks to the contributions of the communicative approach, the focus shifted from discrete-point items to integrative assessment (Larsen-Freeman, 2009) and grammar knowledge is assessed via four skills (Brown and Abeywickrama, 2010). Ozkan (2011) states that using authentic text in teaching and assessing grammar knowledge provides content and face validity. It makes form and meaning assessed equally and improves learners' and testees' motivation. Hughes (2003) suggests that grammar can be used in all kinds of tests and provides content validity.

Brown and Abeywickrama (2010) divide grammar assessment tasks into three classes: selected response tasks, limited production tasks, and extended production tasks. Selected response tasks focus on grammar knowledge. Short sentences are used to check grammar knowledge. Multiple choice tasks, discrimination tasks, noticing task, or consciousness-raising tasks are the techniques for this kind of task. Similar to selected response tasks, limited production tasks require single or longer sentences however they need language production. Since gap filling, short-answer tasks, and dialogue-completion tasks require both form knowledge and comprehension, they are far preferable activities. On the other hand, in extended production tasks, there is authentic language use. Information gap tasks and role-play activities provide authentic language use.

Furthermore, Hughes (2003) proposes gap filling, paraphrasing, completion, and multiple-choice techniques for assessing grammar. Moreover, Heaton (1990) lists the most common grammar testing items as multiple-choice, error-recognition, rearrangement, completion, transformation, changing the words, broken sentences, pairing and matching, combination and addition items. Also, Larsen-Freeman (2009) states that some renovations are proposed to assess grammar knowledge, such as redefining the construct, partial scoring, the social dimension, and the standard.

Finally, a grammar test should include both form and meaning. Testing just accuracy is restricted (Larsen-Freeman, 2013). Text-based grammar tests develop testees' active knowledge of grammar and active corrective feedback increases testees' mastery of grammar (Fotos, 2001).

Vocabulary is the central part of learning a foreign language. A large amount of vocabulary is needed to master a foreign language or communicate in that language. Many people accept learning words as learning a foreign language (Carter, 2001). Vocabulary is a prerequisite for language skills, grammar, and communication.

Since vocabulary is crucial in learning a foreign language, testing this knowledge is crucial, too. Different methods are proposed to test vocabulary. For instance, Read (2000) proposes discrete vs. embedded, selective vs. comprehensive and context- dependent vs. context- independent testing for vocabulary. Discrete tests assess vocabulary knowledge independently from each other and explicitly (Coombe, 2011). This kind of test assesses recognition ability or production ability. The abovementioned recognition ability is that understanding a word whether written or spoken, and production ability is using a word correctly in a written or spoken way (Pignot-

Shahov, 2012). According to Hughes (2003), recognition tests include multiple-choice techniques. In this technique, antonyms, synonyms, definitions, suitable words, or odd words are asked for testees to choose the right answer. On the other hand, tests for production include sentence completion, gap filling, and translation items. On the contrary, embedded vocabulary tests attach importance to the testees' use of the words in their speaking, listening, reading and writing abilities (Coombe, 2011). Another binary is a selective and comprehensive test, which differs in that the former assesses specific words but the latter measures all the words of input or output (Read, 2000).

In the last type of testing vocabulary, context-dependent tests assess testees' vocabulary knowledge in a context to give the appropriate answer. In these tests, testees have clues in the context, yet context-independent tests measure testees' responses without any contextual refference (Read, 2000). Contexts offer real-life situations to use vocabulary. A study conducted by Öztürk (2012) revealed that testees having contextualized tests have better performance than those who have discrete tests. Howbeit, Chun- mei (2007) states that discrete tests provide fast, direct and convincing results in a short time. Moreover, from the communicative point of view, context is not just a linguistic station; it also covers pragmatic knowledge (Brown and Abeywickrama, 2010).Another important distinction of vocabulary test is based on the view of breadth and depth tests. Breadth tests assess the quantity of vocabulary; however, depth tests assess how much the testees know about the word (Laufer and Goldstein, 2004).

As mentioned before, speaking is the main tool of communication. However, for successful oral communication, *pronunciation* is a key factor. Pronunciation in language learning means perceiving and producing sounds in the target language (Seidlhofer, 2001). Testing pronunciation is mostly carried out by testing speaking. Yet, there are some specific techniques for production and recognition levels of pronunciation (Alabbasi, 2007). Alabbasi (2007) classifies production tests into three categories: scored interviews in which each testee is interviewed separately and their results are recorded; highly structured speech samples divided into a few parts and each part assessing different speech samples; paper and pencil tests of pronunciation aiming to check answers pointing pronunciation and stress. The items in these tests are rhyme words, word stress, and phrase stress. On the other hand, recognition tests include sounds and their phonemic symbols, namely, using phonetic alphabet of a foreign language and minimal pairs of words, aiming to distinguish phonemes of pair words, and sound matching with pictures when a tester reads a word, thus testees are expected to remember the phonemes and choose the correct pictures.

Another view for testing pronunciation was proposed by Madsen (1983). He proposes three techniques: limited response, multiple-choice, and reading aloud. The limited response contains individual testing, such as repetition and group testing through identifying the sounds. Despite being easy and testing pronunciation in detail, they are time-consuming. The multiple-choice technique allows testees to choose what they hear or paraphrase by chance. In this technique, context is used rather than isolated items. Besides, testees read a text loud in reading-aloud tests providing a chance to observe the pronunciation. Yet, this technique can reduce reliability due to problems deriving from reading ability (Alabassi, 2007).

In addition, Goodwin (2013) discusses three sorts of pronunciation tests: diagnostic evaluation, ongoing feedback and classroom achievement tests. Diagnostic evaluation is used to determine the learners' needs. During the learning process, ongoing feedback is given to learners. Classroom

achievement tests shed light on the progress.As a result, in pronunciation assessment, there must be a balance between accuracy and intelligibility (McDonough et al., 2013). Intelligibility is the ability to comprehend foreign language articulation (Isaacs, 2014) and including individual sounds, stress, rhythm, intonation, sound, and spelling (McDonough et al., 2013, p.160-161).

Apart from the mentioned components (grammar, vocabulary, and pronunciation) there are other components, such as discourse, semantics and pragmatics. All the language components can be assessed while assessing language skills. Testing the components in isolated situations minimizes the authenticity. Instead, they should be evaluated by means of testing foreign language skills. This kind of evaluation increases authenticity as well as testees' communicative competence.

1.5. Communicative Language Teaching and Communicative Syllabus

Communicative language teaching (CLT) aims to teach a foreign language by means of communication. CLT makes communicative competence the goal of language teaching rather than structuralist competence (Larsen- Freeman and Andersdon, 2011). Communicative competence includes using language for different purposes and functions depending on the settings and participants, producing and comprehending different texts, and using different communicative strategies (Richards, 2006). In other words, foreign language learners need to perform functions, such as promising, inviting, and refusing (Wilkins, 1976). In order to perform these functions, the required syllabus in CLT must contain purposes of communication, settings, roles of learners, communicative events, language functions, notions, the combination of discourse and rhetorical skills, the variety of target language, grammatical and lexical content (van Ek and Alexander, 1980).

As the syllabus type, CLT needs communication rather than abstract rules. This approach proposes a skill-based, functional, notional, and task syllabus (Richards, 2006). In CLT, the teaching is organized according to the functional-notional syllabus (Richards and Rodgers, 1986). Nunan (1988) describes functions as the communicative purposes for the language and notions as the conceptual terms expressed with the language. The well-known and highly accepted functional-notional syllabuses the Council of Europe prepared are the Threshold and Waystage levels (White, 1988). (The threshold level has been called B1 and Waystage A2 since 2001.) As an example, the functions for Waystage level are listed as;

1)   Imparting and seeking factual information
1.1)   Identifying
1.2)   Reporting
1.3)   Correcting
1.4)   Asking
1.5)   Answering questions

2)   Expressing and finding out attitudes
2.1)   Expressing agreement with a statement
2.2)   Expressing disagreement with a statement
2.3)   Enquiring about agreement and disagreement
2.4)   Denying something
2.5)   Stating whether one knows or does not know something or someone
2.6)   Enquiring whether one knows or does not know something or someone
2.7)   Expressing ability and inability

2.8) Enquiring about ability and inability
2.9) Enquiring how (un)certain one is of something
2.10) Enquiring how (un)certain others are of something
2.11) Expressing one is (not) obliged to do something
2.12) Enquiring whether one is obliged to do something
2.13) Giving permission
2.14) Seeking permission
2.15) Stating that permission is withheld
2.16) Expressing want, desire
2.17) Enquiring about want and desire
2.18) Expressing intention
2.19) Enquiring about intention
2.20) Expressing preference
2.21) Expressing pleasure
2.22) Expressing displeasure
2.23) Enquiring about pleasure and displeasure
2.24) Expressing hope
2.25) Expressing satisfaction
2.26) Expressing dissatisfaction
2.27) Enquiring about dissatisfaction
2.28) Expressing disappointment
2.29) Expressing gratitude
2.30) Expressing moral
2.31) Granting forgiveness
2.32) Expressing approval
2.33) Expressing appreciation
2.34) Expressing regret
2.35) Expressing indifference

3)   Getting things done
3.1) Suggesting a course of action
3.2) Requesting others to do something
3.3) Inviting others to do something
3.4) Accepting an offer or invitation
3.5) Declining an offer or invitation
3.6) Enquiring whether and invitation accepted or declined
3.7) Advising others to do something
3.8) Warning others to take care or refrain from doing something
3.9) Offering assistance
3.10) Requesting assistance

4)   Socialising
4.1) Attracting attention
4.2) Greeting people
4.3) When meeting people
4.4) Addressing somebody
4.5) Introducing somebody
4.6) Reacting to being introduced

4.7) Congratulating
4.8) Proposing a toast
4.9) Taking leave

5) Structuring discourse
5.1) Opening
5.2) Hesitating
5.3) Correcting oneself
5.4) Enumerating
5.5) Summing up
5.6) Closing
5.7) Opening (telephone)
5.8) Asking for extension
5.9) Giving notice of a new call
5.10) Opening (letter)
5.11) Closing

6) Communication repair
6.1) Signalling non- understanding
6.2) Asking for overall repetition
6.3) Asking for partial repetition
6.4) Asking for clarification
6.5) Asking for confirmation of understanding
6.6) Asking to spell something
6.7) Asking to write something down
6.8) Expressing ignorance
6.9) Appealing for assistance
6.10) Asking to slow down

<div align="center">(van Ek and Alexander, 1977)</div>

According to White (1988) the selection and grading of the functions are based on some criteria, such as the needs of learners, usefulness, generalization, interest, and coverage. On the other hand, White (1988) states the difficulties in this kind of syllabus as the impossibility of defining function clearly, interpretation, and defining the functions within the context, and not being in isolated contexts.

2. DISCUSSION and CONCLUSION

The present study aims to give overall information about the most preferred test techniques in formal assessment. Testers are still preoccupied with traditional formats (Asassfeh, 2019). Since it is practical, traditional methods are commonly used. However, these techniques are not sufficient enough to give detailed report about neither visual (reading and writing) nor audio (listening and speaking) skills. Although traditional pencil-paper test techniques are highly used, they are perceived as dissatisfied and old-fashioned (Piri, 2022).

Furthermore, it should be noted that many institutions, like The Ministry of National Education of Turkey, run Communicative Language Teaching. CLT aims to use language in real-life situations in order to allow students to use target language real-life situations (CoE, 2001; MEB, 2018). Thus, authentic tasks to assess test-takers knowledge must be taken part in actual performance (Brown, 2007). Boddy and Langham (2000) state that using language in communication is preferable to

using it in isolation so; communicative language tests must be designed and supported. The communicative approach aims to test performance rather than competence (Boddy and Langham, 2000). Canale and Swain (1980) define communicative competence as consisting of grammatical competence, sociolinguistic competence and strategic competence. Since there is discrimination between competence and performance, both use and usage of language must be assessed (Canale and Swain, 1980). According to Phan (2008), communicative tests offer testees creatively using language in communication with all language skills. Moreover, the principles of a communicative test set by Phan (2008) are "start from somewhere" (p. 3), "concentration on content" (p. 3), "bias for best" (p.4) and "working on washback" (p.4).

As a solution, alternative assessment techniques and technology can be fulfilled, as well as traditional ones to create an enriched learning environment (Phongsirikul, 2018). Alternative assessment provides students with authentic use of language in a holistic and integrative way (Hamayan, 1995). Brown and Hudson (1998) mention some benefits of alternative assessment they allow students to display their own performance, focus on both processes and products, show learners' strengths and weaknesses and require higher thinking as Stoynoff (2012) suggests portfolios, projects, conferences, journals, simulations, observations, and interviews can be preferably used.

Besides alternative assessment, technology serves as an effective language skill testing and assessment opportunity (Chapelle, 2010). Particularly, Computer Assisted Language Testing (CALT) is a method that includes not only language testing but also technology assistance (Suvorov and Hegelheimer, 2013). As discussed by Sulaiman and Khan (2019) CALT provides flexible, innovative, personalized, efficient and rapid test practices that fit with the needs of new generation learners. Moreover, CALT can rebuild the testing field by overwhelming the challenges and difficulties of traditional language testing tools (Pathan, 2012).

Traditional written exams are still commonly and mostly preferred. They are practical, reliable, and valid. But in the sense of authenticity they are not useful enough. Yet, alternative assessment and CALT provide authenticity. In this situation instead of choosing either side, all means of assessment can be used. Alternative assessment does not necessarily mean alternative to traditional ones, they can benefit from parallelism (Coombe et al, 2012). Applying principles of only one approach may cause weak assessment in some aspects. As Heaton (1990, p.23) points out "a good test will frequently combine features of the communicative approach, the integrative approach, and even the structuralist approach- depending on the purpose of the test." Testers should be aware of the reality that language is a tool in real life rather than an isolated concept. Language is learned for communicative needs that include all the skills and sub-skills. The communicative approach desires to improve communicative competence. Unless the test items are relevant to the principles of CLT, it results in either the teaching principles and program or the test being ignored. Consequently, the ignored one is the program since the tests force the teachers to teach to test (Wilkins, 1976).

Last but not least, teachers should test to teach, not teach to test. As Volante (2004) specifies reducing the instruction, memorizing, false concepts about school programs, low-validity, and narrow curriculum are the main inappropriate effects of teach to test on teaching process. Testing is supposed to be used to enhance the quality of teaching. It should not be the final stage, but only an indispensable stage of the end-goal, namely teaching/learning language.

In this paper, traditional test items are analysed with respect to Communicative Language Teaching. Further studies that deal with current trends, such as artificial intelligence in testing and assessment or dynamic assessment can be done. Current trends are supposed to be more suitable to communicative principles. Studies with various groups and techniques provide new paradigms in assessment.

REFERENCES

Akıncı, T. (2010). Opinions of English teachers in state primary schools on the test they apply, the effect of SBS on their tests and the problems faced. Unpublished Master's Thesis, Pamukkale University

Alabbasi, A. H. Y. (2007). Techniques of testing pronunciation: Production and recognition levels. *Iraqi Academic Scientific Journals, 3*(5), 14-19.

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

Alderson, J. C. (1996). The testing of reading. In C. Nuttall (Ed.) *Teaching reading skills in a foreign language* (pp.212-228). Heinemann.

Asassfeh, S. M. (2019). EFL Teachers' assessment preferences and prevalent practices: The case of Jordan. *International Journal of Linguistics, Literature and Translation (IJLLT), 2*(7), 67-74

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford University Press.

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

Boddy, N.M. &Langham, C. S. (2000). Communicative language testing: an attainable goal? http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.619.7691&rep=rep1&type=pdf

British Council (2009). https://www.teachingenglish.org.uk/article/testing-assessment/

Brown, S. (2003). Assessment that works at work. *The Newsletter for the Institute of Learning and Teaching in Higher Education*, 11: 6–7.

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy.* (3rd Edition). Pearson Education, Inc.

Brown, H.D. & Abeywickrama, P (2010). *Language assessment: Principles and classroom practices.* (2nd Edition). Pearson Education, Inc.

Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4): 653–675.

Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *JALT Journal, 10*(1 & 2), 15-42.

Buck, G. (2001). *Assessing listening*. Cambridge University Press.

Callison- Burch, C. (2007). *Paraphrasing and translation*. Unpublished Ph.D. Thesis University of Edinburg, Edinburg, Scotland.

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

Carter, R. (2001). Vocabulary. In Ronald Carter and David Nunan (eds.). *The Cambridge Guide to Teaching English to Speakers of Other Languages.* (pp. 42- 47). Cambridge University Press.

Chapele, C. A. (2010). Technology in language testing. http://languagetesting.info/video/main.html

Chun-mei, G. (2007). Influence of contexts on vocabulary testing. *US- China Education Review,* *4*(7), 42-46.

Clay, B. (2001). Is this a trick question? Kansas Curriculum Center.

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

Cohen, A. D. (2001) Second language assessment. In M. Celce- Murcia (Ed.), *Teaching English as a second or foreign language*. (3rd Edition)(pp. 515-535). Heilne &Heilne.

Coombe , C. (2011). Assessing vocabulary in the language classroom. http://marifa.hct.ac.ae/files/2011/07/Assessing-Vocabulary-in-LanguageClassroon.pdf

Coombe, C., Davidson, P., O'Sullivian, P., Stoynoff, S. (2012). *The Cambridge guide to second language ssessment (The Cambridge guides).* Cambridge University Press.

Cunningham, G. K. (1998). Assessment in the classroom: Constructing and interpreting tests. *Australian Journal of Teacher Education*, 23(1).

Demirezen, M. (2013). Testing. In A Sarıçoban (Ed.), *Öğretmenlik alan bilgisi testi İnglizce öğretmenliği* (pp. 165- 194). Murat Yayınları.

Fotos, S. (2001). Cognitive approach to grammar instruction. In M. Celce- Murcia (Ed.), *Teaching English as a second or foreign language*. (3rd Edition) (pp.267-285). Heilne & Heilne.

Gonzalez, A. B. (1996). Teaching English as a foreign language: An overview and some methodological considerations. *RESLA*, 11, 17-49.

Goodwin, J. (2013). Teaching pronunciation. In M. Celce- Murcia, D. M. Brinton & M.A. Snow (Eds.), *Teaching English as a second or foreign language*. (4th Edition) (pp. 134-154).Boston, MA: Heinle ELT.

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 Edition). Pearson Education Limited.

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

Hughes, A. (2003). *Testing for language teachers.* (2nd Edition). Cambridge University Press.

Isaacs, T. (2014). Assessing pronunciation. In A.J. Kunnan (Ed.). *The companion to language assessment* (pp. 140-155). Wiley- Blackwell.

Kitao, S. K. & Kitao, K. (1996a). Testing reading comprehension. ERIC. (ED 398 258).

Kitao, S. K., & Kitao, K. (1996b). Testing speaking. ERIC. (ED 398 261).

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 foreign language.* (4th Edition) (pp. 204-320). Heinle ELT.

Larsen-Freeman, D. (2009). Teaching and testing grammar. In M. H. Long & C. J. Doughty (eds.), *The handbook of language teaching.* Blackwell, 518–542.

Larsen- Freeman, D. &Anderson, M. (2011). *Techniques & principles in language teaching.* Oxford University Press.

Larsen- Freeman, D. (2013). Teaching grammar. In M. Celce- Murcia, D. M. Brinton & M. Snow (Eds*.), Teaching English as a second or foreign language.* (4th Edition) (pp. 256-271). Heinle ELT.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: size, strength and computer adaptiveness. *Language Learning*, 54, 399-436.

Lazaraton, A. (2001). Teaching oral skills. In M. Celce-Murcia (Ed.), *Teaching English as a second foreign language*. Heinle and Heinle.

Lazaraton, A. (2013). Second language speaking. In M. Celce- Murcia, D. M. Brinton & M.A. *Snow (Eds.), Teaching English as a second or foreign language*.(4th Edition) (pp. 106- 120). Heinle ELT.

Madsen, H. S. (1983). *Techniques in testing*. Oxford University Press.

McDonough, J., Shaw, C. & Masuhara, H. (2013). *Materials and methods in ELT*. Blackwell Publishing Ltd.

MEB (Milli Eğitim Bakanlığı). (2018). İlköğretim kurumları (ilkokullar ve ortaokullar) İngilizce dersi öğretim programı. http://ttkb.meb.gov.tr/program2.aspx?islem=1&kno=214

Nunan, D. (1988). Syllabus design. Oxford University Press.

O'Sullivan, B. (2008). Notes on Assessing Speaking. www.lrc.cornell.edu/events/past/2008-

2009/papers08/osull1.pdf

Ozerova, A. (2004). *Types of tests used in English language*. Unpublished Master's Thesis, University of Latvia.

Ozkan, Y. (2011). Assessment of grammatical competence based on authentic texts. *International Journal of English Linguistics, 1*(2), 148-154.

Öztürk, G. (2012). The Effect of context in achievement vocabulary tests. *Journal of Educational and Instructional Studies in the World, 2*(4), 126-134.

Pathan, M. M. (2012). Computer Assisted Language Testing [CALT]: advantages, implications and limitations. *Research Vistas, 1*(4), 30-45.

Phan, S. (2008), Communicative language testing. *TESL Working Paper Series, 6*(2), 1-10.

Phongsirikul, M. (2018), Traditional and alternative assessments in ELT: Students' and teachers' perceptions. *rEFLections, 25*(1), 61-84.

Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language studies working papers*, 4, 37-45.

Piri, B. (2022). *A comparative investigation of English preparatory school students' and instructors' perceptions of using e-portfolio as an alternative assessment tool at university*. Unpublished Master's Thesis.

Read, J. (2000) Assessing vocabulary. Cambridge University Press.

Richard, J. C. & Rodgers, T. S. (1986). *Approaches and methods in language teaching*. Cambridge University Press.

Richards, J. C. (2006). *Communicative language teaching today.* Cambridge University Press.

Riddell, D. (2003). *Teaching English as a second/foreign language.* The McGraw-Hill Companies, Inc.

Saeed, R. R. & Noor, M. (2011) *Handbook for trainers of learning module on test construction techniques*.

Seidlhofer, B. (2001). Pronunciation. In R. Carter and D. Nunan eds., *The Cambridge guide to teaching English to speakers of other languages*, Cambridge University Press: 56-65.

Sulaiman, Z. & Khan, M. (2019). Computer Assisted Language Testing (CALT): Issues and Challenges. 1-11. https://www.researchgate.net/publication/331311015

Suvorov, R., & Hegelheimer, V. (2013). Computer-assisted language testing. In A. J. Kunnan (Ed.), *Companion to language assessment* (pp. 593-613). Wiley- Blackwell.

Van Ek, J.A. and Alexander, L. G. (1977). Waystage. Council of Europe.

Van Ek, J., and Alexander, L. G. (1980). *Threshold level English*. Pergamon

Valette, R. (1977). *Modern language testing*. (Second Edition). Harcourt Brace Jovanovich, Inc.

Volante, L. (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, 35. http://files.eric.ed.gov/fulltext/EJ848235.pdf

Weigle, S. C. (2013). Considerations for teaching second language writing. In M. Celce- Murcia, D. M. Brinton & M. A.Snow (Eds.), *Teaching English as a second or foreign language*. (4th Edition) (pp. 106-120). Heinle ELT.

Weir, J. C. (1993). *Understanding and developing language tests*. Prentice Hall.

White, R. V. (1988). *The ELT curriculum design, innovation and management.* Basil Blackwell Ltd.

Wilkins, D. (1976). *Notional syllabuses.* Oxford University Press.