Testing in English Language Teaching and its Significance in EFL Contexts: A Theoretical Perspective

| Sayyed Rashid Ali* | Hussain Ahmad† | Arshad Ali Khan‡ |

Abstract   *This conceptual paper reviews literature on the most common practices of language tests in ELT contexts around the world. The detailed discussion of various types of language tests is followed by its various aims and objectives which are linked to the qualities or characteristics of different language tests. The review of the literature reveals that language tests and its purposes vary from context to context and there is a wide range of practical constraints that test designers, test-takers, and test administrators encounter. Particularly, in the Saudi EFL context, EFL teachers lack a voice in the process of language assessment and there is a serious dearth of professional development training to raise EFL teachers' awareness of language tests and develop their assessment literacy. The review of the literature suggests that in various EFL/ESL contexts, more top-notch assessment methods need to be introduced and in the Saudi setting.*

Key Words: Assessment, EFL context, ESL teachers, Student proficiency, Language tests.

Introduction

In educational settings, tests are particularly designed to elicit specific behavior which can help us make inferences about various characteristics that individuals may possess. For this purpose, a measuring instrument or tool is designed to obtain a precise sample of an individual's performance or behavior in a given context. The process of obtaining a sample of an individual's behavior involves a systematic gathering of data using specifically designed tools in order to make informed decisions relating to material selection for teaching, teaching strategies, and learners' learning outcomes.

In language teaching and learning contexts, testing plays a pivotal role in assessing language learners' progress in a language classroom. It can also be a tool to determine an individual's performance on achievement tests, such as IELTS and TOEFL. In the field of ELT, four language skills, i.e. reading, writing, listening and speaking are measured using assessment tools that help learners to quantify and see their development as English language learners. Moreover, testing as an integral part of teaching offers ESL teachers significant insights into the achievement and development of English language learners' learning difficulties, learning styles and levels of anxiety (Desheng & Varghese, 2013). These significant insights become the core element of teaching, teaching materials and test formats that facilitate teachers to modify their practices accordingly. In brief, students' scores on proficiency and achievement tests help course designers and test creators to evaluate the appropriateness of teaching materials and the effectiveness of teaching methods for improving and enhancing the knowledge and skills of English language learners.

As the significance of language tests is widely acknowledged, this paper presents a review of the literature on the meaning, purpose, and impact of English language testing on language learners' proficiency. The subsequent sections of this article shows how language testing is practiced in different ESL/EFL contexts and what are the key constituents of an ESL test that best serves the purpose of English language teaching and learning in a typical ESL/EFL context. This paper aims to answer the following three questions:

*Assistant Professor, Faculty of Languages & Translation, King Khalid University, Abha, Saudi Arabia.
†Lecturer, English Language Institute, King Abdul- Aziz University, Jeddah, Saudi Arabia.
‡Assistant Professor, Department of Linguistics and Communication, Lahore, Punjab, Pakistan.
Email: qasimvr@yahoo.com

1. What are the major types of language tests used in EFL/ESL contexts?
2. What are the aims, objectives, and goals of language tests in ELT?
3. How are language tests perceived in various EFL contexts, and particularly in the Saudi EFL context?

The Definition of a Language Test

Theoretically speaking, language assessment is a continuous process that involves a wide array of methodological tools and techniques (Brown & Abeywickrama, 2010). More precisely, "it is about assessing learners' progress, providing them with feedback and deciding on the next step the teaching and learning process" (Sardareh & Saad, 2013, p. 2493). In short, a teacher's appraisal of students' responses in the classroom will be considered assessment. As it is part of daily classroom teaching and learning, good teachers never seize the opportunities to assess their learners' performances. Conversely, a language test is a subset of assessment which consisted of tools that facilitate the collection of quantitative evidence of students' performance at the end of the course. It has many types, as discussed above which are all linked to the purposes of the tests.

In different language contexts, tests and assessments are synonymously used terms; however, both are distinct terms. A test is a tool that measures an individual's ability in a given context. It is a carefully designed instrument that has identifiable and predetermined scoring rubrics. On the contrary, assessment is a continuous process that involves teachers' conscious efforts to keep checking learners' performances throughout the learning period (Brown, 2007).

Teaching and Language Tests

Generally, language teaching and testing go parallel in any language teaching and learning context. Bachman and Palmer (1996) believe that "virtually all language teaching programs involve some testing" (p.8), which assists language teachers to collect and analyze significant information about teaching practices and learners' learning outcomes. Nevertheless, the whole process of testing causes great concerns among different stakeholders, such as students, teachers, test designers, and administrators. Hall (2010) states that "uncertainties inherent in all language test development" (p. 321). Similarly, Hughes (2003) indicates that teachers often "harbor a deep mistrust of tests" mainly owing to their "very poor quality" (p. I). A number of other misconceptions are associated with testing practices, scoring methods, rubrics and unreasonable expectations which often lead to "misunderstanding the nature of language testing" (Bachman & Palmer, 1996, p. 7). Consequently, these flawed and mistaken beliefs and misconceptions often perpetuate issues related to insufficient and inappropriate test contents, unrelated test content to learners' needs, poor test results and feelings of frustration and anxiety (Bachman & Palmer, 1996, p. 6-8).

Types of Language Tests

Literature indicates that language tests have been around for a long time. Although its implementation, design, and content might differ, its purpose in different language teaching and learning contexts remains the same. In Corder's (1973) view:

Language tests are measuring instruments and they are applied to learners, not to the teaching materials or teachers. For this reason, they do not tell us 'directly' about the contribution of the 'teacher' or the 'materials' to the learning process. They are designed to measure the learners 'knowledge of' or 'competence' in the language at a particular moment in his course and nothing else. The knowledge of one pupil may be compared with the knowledge of others or with that of the same pupil at a different time, or with some standard or norm, as in the case of height, weight, temperature etc. (Corder, 1973, p. 351).

Similarly, Halliday et al. (1966) state that "tests are an attempt to construct an instrument for measuring attainment, or progress, or ability in language skills" (p. 215). According to Bachman (1991), "a test is a procedure designed to elicit certain behavior from which one can make inferences about certain characteristics of an individual" (p. 20).

Scholars have divided language tests into two broad categories: a) Testing skills, i.e. reading, writing, listening, and speaking, and subskills, such as vocabulary, spelling, grammar, comprehension, and punctuation;

b) Testing content knowledge. There is a variety of tests such as aptitude test, achievement test, non-referential test and diagnostic test which test language learners' knowledge (Desheng & Varghese, 2013).

The most common types of language tests identified by Alabi and Babatunde (2001) are; achievement, proficiency, aptitude and diagnostic tests. An achievement test is a tool that helps a teacher assess the learning progress of the students by the end of a year. The proficiency test mainly aims to determine a test taker's preparedness for a specific communicative role (Hamp-Lyons, 1998). The proficiency test is a way of predicting learners' future performance and measuring their current proficiency level. Aptitude test is designed to show that a test taker is capable of undertaking a particular job or course of training. It also indicates whether a candidate has developed their ability and interest to learn and improve foreign language proficiency as a result of language teaching instructions. A diagnostic test is a method to determine the strengths and weaknesses of learners in acquiring specific language concepts. The aim of this test is to work on and overcome the learners' weaknesses. Students in a mixed-ability class with a wide range of linguistic backgrounds often have problems with different aspects of language, thus, this test helps teachers to focus on the learners' weaknesses and improve them.

Similar to the types of tests, Malinova and Ivanova (2011) have classified language tests into eight categories. First, tests are linked to the aims of the tests, i.e. aptitude, proficiency, and placement tests. The second category is according to the timing of the tests to understand whether they are allotted limited or unlimited time. The third is based on the administration of the tests, such as group tests and individual tests. Fourth is linked to the different types of tests, for instance, computer-based tests, written tests, and performance tests. The fifth category is according to the types of decisions, such as current tests, preliminary tests, diagnostic tests or final tests. The sixth category is about the format type as test essays and objective tests. Another category is based on the evaluation methods, such as criterion-referenced tests and norm-referenced tests. The last category is in line with the quality of tests, for example, non-standardized and standardized tests.

Ozerova (2004) draws another important distinction between language tests that is the tests are either direct or indirect. In the former, the test aims to check learners' particular skills, such as listening or speaking. Students listen to the tape recording and complete the associated tasks. In the case of the latter, the test aims to measure the language usage of students in a real-life situation, which marks the difference between direct and indirect language tests.

Furthermore, Ozerova (2004) has categorized language tests into subjective and objective tests. As examiner's personal judgment cannot be ruled out in subjective tests, objective tests have no personal judgment at all. This distinction is further elaborated in scoring procedures as examiners make a judgment about the correctness of test-takers' responses based on their subjective understanding of the scoring criteria, whereas the objective tests are marked with a predetermined criteria-a process that involves no judgment of the examiners whatsoever. Oral tests in the form of interviews and written compositions use rating scales that are subjective in nature; however, dictations or cloze tests are the best examples of objective tests (Bachman, 2011).

Similarly, Brophy (2012) categorizes tests into direct and indirect assessments of students. Direct assessment involves direct examination or observation of learners' acquired knowledge or skills that are measured against performance indicators, whereas indirect assessment indicates the self-reports or opinions about the extent or value of language learning experiences (Brophy, 2012).

The Characteristics of a Language Test

As various sorts of language tests show their various goals, they ought to have four qualities to be successful. Initial, a language test ought to be substantial, and it should quantify what it should gauge. Legitimacy has different sorts: face legitimacy, content legitimacy, basis related legitimacy (or prescient legitimacy), factorial legitimacy, develop legitimacy, simultaneous legitimacy, joined and unique legitimacy (Bachman and Palmer, 1996; Hughes, 2003; Flemming and Stevens, 2004; Crocker and Algina, 2006; Bachman, 2011; Brown, and Abeywickrama 2010). Second, it ought to be dependable as the unwavering quality of a test decides the consistency of a test in assessing what it should assess. Dependability demonstrates the degree to which a test, perception, poll or some other estimation apparatus yields similar scores on tedious events (Filiz and Tilfarlioglu, 2017). Third, a language test needs to have objectivity, which would ensure that a test has just one right answer. Fourth, a test ought to be practical and ought not to take a lot of time and vitality. Fifth, a language test must be down to earth

as a dependable and legitimate test may not be helpful on the off chance that it is unreasonable (Bachman, 1990). Viable test "… includes inquiries of economy, simplicity of organization, scoring and elucidation of results" (Bachman, 1990, p. 34). Last, a language test ought to be genuine that would set up a connection between the different highlights of the test and the non-test target-use setting (Bachman, 1990).

Aims and Objectives of Language Tests

The aims and objectives of language tests are identified by various researchers in the field of English Language Teaching (ELT). According to Alabi and Babatunde (2001), language test determines students' learning outcomes based on the prescribed syllabus. They believe that the language test is a diagnostic tool to assess learners' strengths and weaknesses. Moreover, it helps the teachers to develop familiarity with different types of tests, such as aptitude and proficiency tests.

Language tests have pedagogical objectives to achieve (Bachman & Palmer, 1996). In Buck's (2001) view, language tests aim to offer learning opportunities for students and administrators. Although tests may serve different purposes, they need to reliable, fair and valid, which will allow the administrators to learn about the test-takers' communicative competence. In the same way, Tomlinson (2005) considers language tests as a means of new learning opportunities that enable the language to acquire topnotch knowledge and enhance skills and awareness in the process of taking different tests.

In English as a foreign language (EFL) context, Ozerava (2004) linked the types of tests to their aims. For instance, diagnostic tests are to enable language teachers to present the learners' performances in the form of a diagram, both at the start and at the end of the year to compare the learners' results and show if they have made progress in their final tests. Similarly, a placement test is designed and administered to streamline language learners into appropriate groups based on their proficiency levels. Another type of test in which a proficiency test aims to check the learners' level of communicative competence by the end of the studies. The last one is the progress test that helps language teachers to assess students' progress and see whether the taught materials were successfully learned by the learners.

Language Testing in ELT

The above sections have discussed the types and characteristics of language tests in the orbit of ELT. Acknowledging the significance of language tests, researchers in the EFL/ESL world have conducted a plethora of studies to show empirical evidence of the nature, status, and role of language testing in different contexts.

Ridhwan (2017) conducted a library study on an understanding of formative and summative assessment learning in which he reviewed relevant studies, identified issues, and proposed solutions. In his meta-analysis of the literature, he highlights the point that language teachers have concerns about the nature of assessment tools used in classrooms, which allow learners to cheat while classroom assessment. Ridhwan (2017) suggests that teachers should not merely rely on tests to measure learning outcomes; rather, they can use other formative procedures to give feedback on learners' performance and help them achieve the learning targets. This approach might enable the teachers to deal with problems, such as cheating in exams or lower students' grades and transform their practices from being an assessment of learning to assessment for learning.

In the Korean context, Sook (2003) identified different types of speaking assessment tasks that were used by junior secondary school English teachers. The qualitative responses of the EFL teachers on the study pointed out various practical constraints that affected the assessment of oral skills. The findings indicated that unlike other tests, the speaking tests were psychologically less challenging. They were time-saving and designed specifically for the convenience of administration and construction. Interestingly, the Korean language teachers were not worried about the reliability and validity of their testing tools as they had little or no theoretical knowledge of the speaking assessment, due to which the teachers lacked confidence in carrying out speaking assessment. As the findings suggested practical constraints encountered by the Korean EFL teachers, these include; time-consuming tests due to large classes that led to excessive workload in addition to face-to-face teaching, lack of efficient and effective testing tools, lack of knowledge and training in conducting speaking exams and finding it difficult to elicit learners' responses on the speaking test.

In an identical study, Sancha (2007) investigated Capeverdean teachers' perceptions of using communicative

testing tools to effectively assess speaking competence of EFL learners. The qualitative data indicated that Capeverdean teachers faced various practical issues in the process of communicative testing of EFL students. The findings suggest that language teachers should adopt innovative ways of evaluating learners' oral proficiency in order to avoid practical constraints, mainly associated with the implementation and administration of the test.

Rogers, Cheng, and Hu (2007) carried out research to investigate EFL teachers' beliefs about language testing, assessment and evaluation. This study is interesting in the sense that it includes EFL teachers from Canada, Hong Kong, and Beijing. Interestingly, the quantitative results showed great similarities among instructors from three different contexts. The study confirms the fact that assessment and evaluation were significant factors in terms of improving instructional practices and developing learners' proficiency. More importantly, the results suggest that assessment and evaluation methods varied from context to context and it was based on the teachers' understanding of their assessment practices that determine their implementation or preparation of assessment tools. The findings revealed that assessment has two major categories; a) administrative functions and implementation of procedures using pen and paper, and b) performance assessments which are more natural than paper and pencil assessment. Similar to others' findings, this study also found a difference in the instructors' preparation, training, and confidence in administering or applying evaluation tools. In a similar study, Cheng et al., (2004) reported identical beliefs of EFL/ESL teachers in three different contexts, which were specific to the context.

Moving from teachers to students, Vavla and Gokaj (2013) conducted a study to understand learners' perceptions of assessment and testing in the Albanian EFL context. Their work mainly underlined the effects of language tests on learners and the pros and cons of language tests in an EFL context. Following a mixed-method approach, the findings showed that despite seeing tests as a source of demotivation, they are mandatory and essential to assess and examine learners' outcomes. Although students had no voice in language tests which were solely prepared by teachers, students were in a position to reflect on different aspects of the tests and recommended changes in the structure and administration of tests.

More on EFL learners' experience of testing, Tütünis (2011) conducted a small-scale quantitative study to investigate the washback effects of testing on EFL learners' writing skills. Their results are indicative of learners' anxiety on writing tests. The main sources of anxiety are lack of knowledge about words, grammar and syntax; fear of failure, time constraints, fear of negative evaluation.

In the Iranian EFL context, Hashemi, Khodadadi, and Yazdnmehr (2009) conducted a mixed-method study on the EFL learners' writing tasks to understand the effectiveness of (ESOL English for Speakers of Other Languages) exam preparation courses, such as IELTS, TOEFL, FCE and CAE. The authors believed that EFL writing tests could be evaluated by looking at the general features of appropriateness and the procedures for undertaking the writing tasks. As the students or participants were put into two groups, the aim was to prepare one group for the actual test, whereas the second group had only a goal to develop their proficiency in general. The samples of students' writing tasks of both the groups were collected and their ratings were compared. The second group performed reasonably well as they had no pressure of time meeting certain testing criteria. Based on the find in the gs, the study suggested various techniques that could help teachers improve their practice of teaching writing and preparing students for ESOL examinations.

Language Testing in the Saudi EFL Context

There isn't a lot of research on testing and assessment did in the Saudi EFL setting. The examinations referred to in this segment are the ones that feature the way that more research on this issue should take the analysts' time and exertion to improve the testing strategies and methodology and yield increasingly powerful outcomes. Al Sadaawi (2010) is of the view that prior Saudi instruction framework embraced various quantitative instructive techniques to improve the nature of learning results; nonetheless, more as of late, there is a move towards progressively subjective strategies. This move for the most part because of the acknowledgment that Saudi instructive procedures are not the universal gauges. There has been an expanding number of papers that have scrutinized the current educational programs, nature of showing philosophies, understudies' exhibitions, and assessment techniques. Al Sadaawi (2010) accepts that the developing requests for the institutionalization of evaluation strategies over King have been connected to improved learning results of the understudies and a

compulsory piece of instructive changes.

There have been a couple of studies that have inspected the adequacy of evaluation techniques, at present applied in various Saudi schools and colleges. Obeid (2007) directed a little scale exploratory investigation to comprehend EFL instructors' and understudies' impression of composing assessment the at college level. The discoveries of the investigation featured a few concerns raised by the instructors and understudies. Like the Korean setting discussed the in above area, Saudi educators too found commonsense imperatives while instructing and evaluating composing aptitudes. Instructors considered the appraisal apparatuses improper and deficient to assess the students' composing contents, while, the EFL students accepted that the evaluation techniques were invalid as they would regularly differ with the given evaluations and proclaim them out of line. Thusly, EFL educators think that it's hard to persuade or fulfill language understudies who have exclusive standards from instructors and language tests.

In another examination, Umer, Zakaria, and Alshara (2018) attempted to demonstrate the point that Teacher AssessmLiteracyacry (TAL) can affect EFL students' learning result, and the key explanation behind EFL instructors' inability to lead successful language tests is their absence of hypothetical comprehension of language testing. They found that Saudi EFL instructors in the advanced education needed information and competence to carry out your EFL evaluation in an institutionalized manner. In a blended technique study, they examined the EFL instructors' development of appraisal assignments, the effect of the errands on students' results, and the educators' capacity to cond the cut evaluation as per the suggested practices. The discoveries demonstrated inconsistencies between understudies' learning results and educators' appraisal undertakings. Moreover, the appraisal undertakings, for the most part, started decay learning and retention of learning procedures. The hole between appraisal undertakings and learning results recommend that EFL instructors ought to devel hack more noteworthy comprehension of how to actualize evaluation apparatuses and in this manner, the expert advancement courses require to concentrate on creating appraisal education of language educators, which will improve the learning results of Saudi EFL learners at the national level.

The fascinating point about the examination by Umer et al., (2018) is that they have presented the subject of educator proficiency and information on appraisal, recognizing and underlining the way that EFL instructors in the Saudi setting need to build up their language evaluation education to apply standards of sound evaluation. The thought of TAL isn't novel as researchers around the globe have remembered it as a major proficient learning necessity in cutting edge instructive frameworks around the globe (Volante and Fazio, 2007; DeLuca 2012; Popham 2013; DeLuca, LaPointe-McEwan, and Luhanga, 2016). Besides, there is a far-reaching understanding among the researchers that appraisal proficiency can improve language educators' capacity to create and actualize great quality evaluation devices (Stiggins, 2002, 2004; Popham, 2004; Plake, 2015).

More on the Saudi EFL educators' recognitions and convictions about their job during the time spent language evaluation, Mansory (2016) did a doctoral project in a Saudi Arabian college. His two-dimensional examination explored not just the EFL educators' job is persistent and summative evaluation rehearses, however it likewise revealed EFL instructors' understanding and demeanor towards appraisal rehearses in the Saudi EFL setting. The subjective discoveries of this huge scale study uncovered that EFL educators had no job in summative appraisal at all. Just those bunch of instructors who were individuals from the appraisal advisory group may have a state in the structure or execution of the tests; be that as it may, educators everywhere were disregarded in such manner, was were liable for directing or reviewing the tests while following the recommended methodology. The discoveries uncovered a solid interest from the language instructors to be effectively engaged with the procedure of language evaluation. Be that as it may, because of top-down initiative structures and approaches, the administration didn't permit EFL instructors to have their state in the appraisal procedure as educators frequently condemn the current arrangements and methods.

In an attempt to see the appropriateness and effectiveness of EFL tests, Alfallaj and Al-Ahdal (2017) conducted a small-scale study. Their pilot study indicated that the design of the question papers lacked consistency. They were also not in line with modern assessment techniques and lacked experts' input. Therefore, the study suggested that test designers and teachers should be trained to meet the international standards, which would also have bearings on curriculum and pedagogy.

As language test anxiety is a widely researched topic, in the Saudi context too, it has received some attention.

Alrabai (2004) believes that feeling of anxiety often affects Saudi EFL learners' learning outcomes. In a large-scale study with a mixed-method approach, Alrabai (2004) investigated factors that led to the Saudi EFL learners' foreign language anxiety (FLA). The findings revealed that learners often felt anxious and did not perform up to their levels.

Alotabi (2014) investigated the major techniques utilized by Saudi EFL teachers for assessing the students' language skills. This was a general study not focused on language tests; rather, its main aim was to understand the role of teacher and student inside the classroom valuations. Although the study does not clearly indicate whether Alotabi (2014) is concerned with formative or summative assessment, assessment techniques and student-teacher interaction were generally investigated in order to understand the assessment practices. According to Alotabi (2014), as teachers ask spontaneous questions and force students to think and respond. This study is different from all other traditional studies on testing, evaluation or assessment as it has used student-teacher interaction and teacher questioning techniques as a tool of assessment in a language classroom. Although the findings do not offer help in understanding the effectiveness and the impact of the current technique on learners' performances, it can be further considered as a potential area of research for the researchers in the field of ELT.

Conclusion and Further Research

This conceptual article reviews the literature on the notion of testing and assessment in the field of ELT. The definition of a language test is not specific to a context; rather, it is a yardstick that is used to measure EFL/ESL learners' achievement, performance or progress in a given context. although language tests have different kinds, the most common ones are aptitude tests, achievement tests, progress tests, diagnostic test and proficiency tests. These different tests have different purposes in a language teaching and learning context that influence its contents, implementation, and administrative procedures. The types and aims of language tests may vary; however, their characteristics, such as reliability, validity, objectivity and practicality remain unchanged.

The article has also reviewed studies on language testing in different EFL contexts. The conclusions of different studies have shown great similarities in irrespective of their contexts. For example, EFL teachers lack knowledge and training in conducting language tests, particularly, the speaking test. Grading writing script is another problem faced by the teachers. Overall, the review of the literature has pointed out practical constraints, administrative hurdles, lack of voice and lack of training in the process of designing and implementation of language tests.

As most of the studies show EFL/ESL teachers' perceptions or experiences of language tests, fewer studies have included EFL/ESL learners' views. As students are mostly influenced by a language test, future research should take their perspective into account, so the test designers or planner could make changes accordingly. Although studies have investigated EFL/ESL teachers' perceptions. More importantly, researchers have mainly employed quantitative tools to gather data, however, the combination of qualitative tools can give useful insights into learners' experiences and issues related to language tests.

References

Alabi, A.O. & Babatunde, M.A. (2001). *Primary English curriculum and methods*. Oyo, Odumatt Press and Publishers.

Alfallaj, F. S. S., & Al-Ahdal, A. A. M. H. (2017). Authentic assessment: Evaluating the Saudi EFL tertiary examination system. *Theory and Practice in Language Studies*, 7(8), 597 607.

Alotabi, K. A. (2014). Student assessment strategies in Saudi Arabia: a case study of pre and post classroom practices. *Literacy Information and Computer Education Journal (LICEJ)*, 3(1), 1786-1763.

Alrabai, F. (2014). A model of foreign language anxiety in the Saudi EFL context. *English language teaching*, 7(7), 82.

Alsadaawi, A. S. (2017). Factor Structure and Reliability of Test Items for Saudi Teacher Licence Assessment. *International Education Studies*, 10(2), 26.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L. F. (1991). What does language testing have to offer?. *Tesol Quarterly*, 25(4), 671-704.

Bachman, L. F. (2011). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Brophy, T.S. (2012). *Writing effective rubrics*. Institutional Assessment Continuous Quality Enhancement Series.

Brown, D., & Abeywickrama, P. (2010). *Language assessment principles and classroom practices (2nd ed.)*. Pearson Education.

Brown, H.D. (2007). *Teaching by principles. "An interactive approach to language pedagogy"* Pearson, Longman

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

Corder, S. P. (1973). The elicitation of interlanguage. *Errata: Papers in error analysis*, 36-48.

Crocker, L., & Algina J. (2006). *Introduction to classical and modern test theory*. Thomson Learning.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. Assessment in Education: Principles, Policy & Practice, 17, 419-438.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Teacher assessment literacy: a review of international standards and measures. Educational Assessment, Evaluation and   Accountability, 28(3), 251-272. Retrieved from https://eric.ed.gov/?id=EJ1112727

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Teacher assessment literacy: a review of international standards and measures. Educational Assessment, Evaluation and   Accountability, 28(3), 251-272. Retrieved from https://eric.ed.gov/?id=EJ1112727

Desheng, C., & Varghese, A. (2013). Testing and evaluation of language skills. *IOSR Journal of Research & Method in Education.*, 1(2), 31-33.

Flemming, M. & Stevens, D. (2004). *English Teaching in the Secondary School*.(2nded.), 113 114). London: David Fulton Publishers Ltd.

Hall, G. (2010). International English language testing: a critical response. *ELT Journal*, 64(3): 321–328.

Halliday, M.K.A. et al. (1966). *The Linguistics Science and Language Teaching*. London: Longman

HAMP-LYONS, L. I. Z. (1998). Ethical test preparation practice: The case of the TOEFL. *TesolQuarterly*, 32(2), 329-337.

Hashemi, M. R., Khodadady, E., & Yazdanmehr, E. (2010). Learners evaluation of EFL writing tasks in Iran ESOL exam preparation courses. *Journal of English Language Teaching and Learning*, 52.

Hughes, A. (2003). *Testing for language teachers.* Cambridge: Cambridge University Press.

Malinova, А., & Ivanova, V. (2011). Automation of Electronic Testing Examination in English Language using DeTC. *Plovdiv University "Paisii Hilendarski" Scientific Works*, 38, 59-68.

Mansory, M. (2016). EFL Teachers' Beliefs and Attitudes towards English Language Assessment in a Saudi University's English Language Institute.

Obeid, R. (2017). Second Language Writing and Assessment: Voices from within the Saudi EFLContext. *English Language Teaching*, 10(6), 174-181.

Ozerova, A. (2004), *Types of Tests Used in English Language.* Bachelor Paper. S. (2, 4, 8, 9, 10, 11, 12, 13, 14, 15, 16)

Plack, B. (2015). Assessment in the Service of Learning. Presented in the Second International Conference for assessment and evaluation: Learning outcomes assessment. Retrieved from http://ica. qiyas. sa/ Presentations/ ResearchSummary/En.pdf

Popham, W. J. (2004). Why assessment illiteracy is professional suicide. Educational Leadership, 62, 82-83. Retrieved from http://www.ascd.org/publications/educationalleadership/sept04/vol62/num01 /Why-Assessment-Illiteracy-Is -Professional-Suicide.aspx Popham, W. J. (2013). *Classroom assessment: What teachers need to know* (7th ed.). Boston: Pearson.

Ridhwan, M. (2017). Understanding Formative and Summative Assessment for EFL Teachers:nTheoretical Reflections on Assessment for Learning. *Journal of English for Academic*, *4*(1), 40-50.

Rogers, W. T., Cheng, L., & Hu, H. (2007). ESL/EFL instructors' beliefs about assessment andevaluation. *Comparative and International Education/Éducation Comparée et Internationale*, *36*(1), 39-61.

Sancha, M. A. D. C. L. (2007). *Communicative EFL Assessment in Cape Verde: Perceptions, Constraints and Suggestions for Effective Testing Tools* (Bachelor's thesis).

Sardareh, S. A., Saad, R, M. (2013). Defining assessment for learning: A proposed definition from a sociocultural perspective, Life Science Journal, 10(2), 2493- 2497

Sook, K. H. (2003). The types of speaking assessment tasks used by Korean Junior Secondary school English teachers. *Asian EFL Journal*, 1-11.

Stiggins, R. (2002). Assessment crisis: the absence of assessment for learning. Phi Delta Kappan, 83, 758-765. https://doi.org/10.1177/003172170208301010

Stiggins, R. (2004). New assessment beliefs for a new school mission. Phi Delta Kappan, 86, 22-27. https://doi.org/10.1177/003172170408600106

Tilfarlioglu, F. Y. (2017). Testing and evaluation in ELT methodology. Retrieved from https://www.researchgate.net

Tomlinson, B. (2005). *Testing to learn: a personal view of language testing*. ELT Journal, Oxford University Press 59/1, 44

Tütüniş, E. Ş. B. (2011). The washback effect of testing on students' learning in EFL writing classes. http://dergiler.ankara.edu.tr/dergiler/27/1831/19272.pdf

Umer, M., Zakaria, M. H., & Alshara, M. A. (2018). Investigating Saudi University EFL Teachers' Assessment Literacy: Theory and Practice. *International Journal of EnglishLinguistics*, *8*(3), 345.

Vavla, L., & Gokaj, R. (2013). Learner's Perceptions of Assessment and Testing in EFL   Classrooms   in Albania. *Mediterranean Journal of Social Sciences*, *4*(11), 509.

Volante, L., & Melahn, C. (2005). Promoting assessment literacy in teachers: Lessons from the Hawaii school assessment liaison program. Pacific Educational Journal, 13, 19-34.