

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/321155250>

Testing and Evaluating in ELT Methodology

Chapter · January 2017

CITATIONS

0

READS

11,788

1 author:



Filiz YALCIN Tilfarlioglu

Gaziantep University

25 PUBLICATIONS 142 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Integration of Lean method in English Language Teaching [View project](#)



Lean-centered model [View project](#)

TESTING AND EVALUATION IN ELT METHODOLOGY

Filiz YALÇIN TILFARLIOĞLU

A golden key can open any door...

INTRODUCTION

This chapter summarizes two golden keys in ELT methodology: **testing and evaluation**. In ELT methodology, the term of ‘test’ can be defined by using different perspectives. There are various definitions of test in the literature in classroom research. Tests are any instruments of evaluation which measure ability, knowledge or performance by eliciting observable behavior from the test taker (Richards, 1999; Allan, 1995; Genesee & Apshur, 1996) as they try to make accurate predictions about small samples of performance in a complex structure like language (Allan, 1995). In language context, a ‘test’ is designed to measure and evaluate students’ language proficiency according to different qualities (Hughes, 2003). Bachman and Palmer (1996) mention the benefits of tests by stating that a test is useful when it has authenticity, interactiveness and practicality in addition to measurement qualities which are reliability and validity.

Reliability:

An important aspect of test is being reliable. *Reliability* is defined as the extent to which a questionnaire, test, observation or any measurement tool produces the same results on repeated trials. In short, it is the stability or consistency of scores over time or across raters. Similarly, the degree to which an individual’s responses (i.e., their scores) on a survey would stay the same over time is also a sign of reliability. There are three aspects of reliability, namely: equivalence, stability and internal consistency (Miller, 2005).

Validity:

Validity is also an important term related to tests and it is defined as the extent to which the instrument measures what it purports to measure. For example, in a test that is used to test students’ listening ability, grammar should not be scored to preserve validity. There are many different types of validity: content validity, face validity, criterion-related validity (or predictive validity), construct validity, factorial validity, concurrent validity, convergent validity and divergent (Hughes, 2003; Bachman, 2011; Bachman & Palmer, 1996; Brown,

2010; Crocker & Algina, 2006; Flemming & Stevens, 2004). According to Fowler (2002), there are six guidelines to review the EFL exam for content validity criteria: (1) clarity in wording, (2) relevance of the items, (3) use of Standard English, (4) Absence of biased words and phrases, (5) formatting of items, and (6) clarity of the instructions. It should be kept in mind that reliability is necessary but not sufficient for validity. That is, for something to be valid it must be reliable but it must also measure what it is intended to measure (Miller, 2005). According to Crocker and Algina (2006), two major questions have to be answered when discussing the validity of an instrument:

- Is the scale measuring the construct intending to be measured?
- Is there sufficient evidence to support the intended uses or interpretations of the test?

Pedagogical Purposes in Testing:

Tests can serve pedagogical purposes (Bachman & Palmer, 1996) and can be classified differently in terms of various criteria. Buck (2001) claims that

“The main purpose of language testing is to provide opportunities for learning, both for the students who are being tested, and for the professionals who are administering the tests. It is obviously important that test should be fair, valid, and reliable, the most important of all is that tests should provide useful opportunities for learning. A good test is one that allows us to make useful inferences about the test taker’s communicative language ability”(p:92).

Similarly, Tomlinson (2005) also contributed the issue and supports that “Tests are not normally viewed as events in which new learning takes place. But it is perfectly possible for learners to gain new knowledge and to develop new awareness and skills whilst actually taking a test”.

What are Test Categories?

Related to this, Ivanova (2011) divided tests into eight basic categories: according to *purpose* (proficiency, placement, and aptitude tests); according to *test timing* (whether limited time or unlimited), according to *test administration* (individual tests and group tests). The fourth category is *answer type* (written tests, computer tests, and performance tests). The fifth is according to *decision type* (preliminary tests, current tests, final tests and diagnostic tests). Another category is according to *item format type* (objective tests and tests essays). The seventh category is according to *evaluation method* (norm-referenced tests and criterion-

referenced tests). Final category is according to *test quality* (standardized tests and non-standardized tests).

Types of Testing:

When the aims of tests are considered, Anjelika Ozerava (2004) asserts that there are many testing types which are used for different purposes in EFL context. One of them is the *diagnostic test*, the teachers are supposed to analyze the diagnostic tests by completing special documents and by providing diagrams with the results of each class. Then, at the end of the study year the teachers are demanded to compare the results with the final tests. Another type is the *placement test* which is used to decide which group or class the learner could be joined to. It assists to put the student exactly in that group that responds his/her true abilities. The *progress test* is used to show the teacher whether the students have learnt the recently taught material successfully. The last one is *proficiency test* which intends to check the learners' language competence.

According to Ozerova (2004), tests differ whether they are *direct* or *indirect*. In *direct test*, teacher is interested in testing a particular skill. For instance, if the aim of the test is to check students' listening skills, teacher uses the direct test such as listening to the tape and doing the accompanying tasks. On the other hand in *indirect testing*, tests measure the usage of the language in real-life situation. The appeal to 'real-life performance' provides the basis for a distinction between 'direct' and 'indirect' tests of proficiency. According to Clark (1975), in direct proficiency testing, the testing format and procedure attempts to duplicate as closely as possible the setting and operation of the real-life situations in which the proficiency is normally demonstrated. Clark (1978) also pointed out that indirect measures are not required to reflect authentic language-use contexts and, indeed, they may in many cases bear little formal resemblance to linguistic situations that the student would encounter in real life.

Testing types also differ depending on our purpose. The techniques that are used for testing change accordingly. There are many types of tests with different purposes. Teacher should be aware of the purposes by using these tests because sometimes testing can have disastrous effects on student motivation. "Testing type can entirely change the students' attitude towards learning the language" (Ozerova, 2004). Teachers have a variety of assessment alternatives from which to select as they focus on the valued learning targets.

Accurate assessment conclusions are dependent on the selection or development of proper assessment tools. The options include selected response (multiple choice, true/false, matching and fill in), extended written response, performance assessments (based on observation and judgment), and direct personal communication with the student. The challenge in all contexts is to match an assessment method with an intended achievement target (Chappuis, 2004).

As an example, Tomlinson (2005) purports that major *summative examinations* must be reliable and fair, and must discriminate between the candidates; therefore their means of assessment are still predominantly objective, and they still rely to a large extent on such task types as multiple choice, transformation, substitution, and matching. The researcher also mentions that *formative tests* which give information about a learner's progress during a course. The researcher states that they provide useful opportunities for learning.

Another way of testing is *the norm-referenced testing* which measures the knowledge of the learner and compares it with the knowledge of another member of the group. Norm-referenced tests are designed to enable the test user to make 'normative' interpretations of test results. That is, test results are interpreted with reference to the performance of a given group, or norm. The 'norm group' is typically a large group of individuals who are similar to the individuals for whom the test is designed. In the development of norm-referenced tests, the norm group is given the test, and then the characteristics of this group's performance are used as reference points for interpreting the performance of other students who take the test. Likewise, *criterion-referenced tests* are designed to enable the test user to interpret a test score with reference to a criterion level of ability or domain of content. An example would be the case in which students are evaluated in terms of their relative degree of mastery of course content, rather than with respect to their relative ranking in the class. Thus, all students who master the course content might receive an 'A', irrespective of how many students achieve this grade. The primary concerns in developing a *criterion-referenced test* are that it adequately represent the criterion ability level or sample of the content domain, and that it should be sensitive to levels of ability or degrees of mastery of the different components of that domain (Bachman, 2011, p. 72-74).

Ozerova (2004) also mentions the objective and subjective tests. In objective tests, there is no judgment, but subjective tests involve personal judgment of the examiner. As Pilliner (1968) pointed out, 'subjective' tests are distinguished from 'objective' tests entirely in terms of scoring procedure. In an objective test the correctness of the test taker's response

is determined entirely by predetermined criteria so that no judgment is required on the part of scorers. In a subjective test, on the other hand, the scorer must make a judgment about the correctness of the response based on her subjective interpretation of the scoring criteria. The multiple-choice technique is the most obvious example of an objective test, although other tests can be scored objectively as well. Cloze tests and dictations, for example, can be scored objectively by providing scorers with scoring keys that specify exactly which words are acceptable and which are not. Tests such as the oral interview or the written composition that involve the use of rating scales are necessarily subjectively scored, since there is no feasible way to ‘objectify’ the scoring procedure. (Bachman, 2011, p.76)

How to Form a Test?

Forming tests is another issue which the researchers focused in throughout the literature. Wolfe (2004) is one of these researchers who emphasizes that test questions can take several forms. Selections questions, such as true/false, multiple choice and matching can be graded easily and take less time for students to answer. Multiple choice exams are designed for the recall of facts and an understanding of concepts. In a gap-fill or cloze test item, the student is asked to select the term that goes into it from a small number of candidate for each gaps presented in the text. Multiple-choice cloze test is useful for proficiency testing (Simith, Avinesh, & Kilgariff, 2010). Jumbled sentences can be used when practicing correct word order. There are also assignment tests which are used when teachers assign the task to students. They send for example their translations of original texts by this way. We hold on-line courses in high regard for their possibility to control students’ work and their study progress continuously through controlling and testing tools. The teacher can use the tests also for final testing of students, if only he excludes any possibility of cheating (Hubackova, Semradova & Klimova, 2011).

In ELT context, four skills require different procedures in a testing process. For example, in an oral exam, creating an authentic atmosphere and natural conversation is challenging. Underhill (1987) states that students must be relaxed and confident in order to communicate, therefore “allowing the activity (a conversation) to become dominant, and its ulterior purpose (a language test) to be temporarily subordinated. The oral test then reaches its highest degree of authenticity by no longer being a test”. In an oral test, different techniques can be used such as oral report, learner-learner joint discussion/decision making, role-play, interview, question and answer, using a picture or picture story, sentence completion from

aural or written stimulus, sentence correction etc. Underhill (1987) mentions that, interview is the most common of all oral tests. It is direct and face-to-face exchange between student and the interlocutor. In its traditional form, however, it has at least one potentially serious drawback. The relationship between the tester and the candidate is usually such that the candidate speaks as to a superior and is unwilling to take the initiative. As a result, only one style of speech is elicited, and many functions (such as asking for information) are not represented in the candidate's performance. It is possible, however, to get round this problem by introducing a variety of elicitation techniques into the interview situation (Hughes, 2003). Similarly, in role-play technique, student takes a role in a particular situation and is expected to converse with the interviewer in a way that is appropriate to the role and the situation given (Underhill, 1987). Reinmann (2003) uses role-play in his study in order to use communicative means to expose the students to practical and authentic language, which they could practice and use appropriately, within real life situations. "This means of testing was used because it provided a way to replicate the real life qualities of language and other non-linguistic factors, which are necessary for successful communication" (Reimann, 2003).

EVALUATION & ASSESSMENT

Assessment:

Assessment and evaluation are essential components of teaching and learning in English language arts. Without an effective evaluation program it is impossible to know whether students have learned, whether teaching has been effective, or how best to address student learning needs. The quality of the assessment and evaluation in the educational process has a profound and well-established link to student performance (Foundation for the Atlantic Canada English Language Arts Curriculum).

Assessment (Hanna & Dettmer, 2004) is the process of gathering data. More specifically, assessment is the ways instructors gather data about their teaching and their students' learning. Assessment, is an ongoing process that measures progress for example it is not limited to the final achievement or summative tests at the end of the term, and it can measure learners' learning along quite diverse ways; for instance, when a learner offers a comment or responds to a question, the teacher subconsciously makes an estimate of that learner's state of knowledge. Assessment is important to all participants of learning process who are learners, their parents, their teachers and the school as a whole (Medwell, Wray Minns, Coates, & Griffiths, 2009; Khonbi & Sadeghi, 2013). Assessment of progress also

enables teacher to evaluate teaching methods, approaches and materials that are used. The extent of students' learning is an important and necessary criterion for judging the success of teaching approaches (Medwell, Wray, Minns, Coates & Griffiths, 2009). Black, Harrison, Lee, Marshall, and Wiliam (2004) state that when an assessment activity provides information to be used as feedback, it can enforce learning and can help to modify the teaching and learning activities in which they are engaged.

According to Chan (2007):

“No single assessment is able to thoroughly cover the learning progress or achievement of students. The notion of “multiple” in multiple assessment can be interpreted from different administrative perspectives : (a) formative assessment is administered during the learning process with the aim of using the results to improve instruction; (b) summative assessment is performed at the end of a course often for purposes of providing aggregate information on program outcomes to educational authorities; (c) traditional test- based assessment involving written or oral tests; and (d) task-based assessment which are activity-based, and includes teacher-student interaction, student- student interaction, body movement, teamwork, portfolios, and other language requirement and activities”.

On the other hand, formative assessment provides feedback and information during the instructional process, while learning is taking place, and while learning is occurring. It measures student progress but it can also assess your own progress as an instructor. Types of formative assessment: observations during in-class activities, question and answer sessions, in-class activities where the students informally present their results, conferences between the instructor and student at various points in the semester. Black and William (1998) state that, growing in popularity, formative assessment is to a greater extent promoted as ‘a moment of learning’. It allows beneficial comparisons between the actual and referral levels of achievement, and so the results can be used to identify gaps between the compared levels. Huges (2003) emphasizes that:

“Testing is not, of course, the only way in which information about people’s language ability can be gathered. It is just one form of assessment, and other methods will often be more appropriate....Assessment is formative when teacher use it to check on the progress of their students, to see how far they have mastered what they should have learned, and then use this information to modify their future teaching plans. Such assessment can also be the basis for feedback to the assessment but so will simple observation (of performance on learning tasks, for example) and the study of portfolios that students have made of their work. Students themselves may be encouraged to carry on self-assessment in order to monitor their progress, and then modify their own language learning objectives” (p. 5).

Summative assessment (examinations, term papers, projects, portfolios, performances, final examination, student evaluation of the course, instructor self-evaluation) takes place after the learning has been completed and provides information and feedback that sums up the teaching and learning process.

Chan (2007) contributed to the issue as well and explains that

“Assessment can be multiple, comprising vocabulary, sentence patterns, songs and rhymes as well as the skills of listening, speaking, reading, and writing. Assessment can also be multiple in terms of tools. It can involve the traditional paper-and-pencil test or use of multiple media such as computer, tape recorder, or video recorder. In terms of the tester, assessment can be multiple, engaging teachers, peers and self (Brown, 2001). Moreover, in test-based assessment, discrete-point tests which focus on linguistic forms and structures and integrative tests which emphasize comprehension and application of language can be used based on the content and purpose of the assessment”.

According to Brophy (2012) “Direct assessments of student learning are those that provide for direct examination or observation of student knowledge or skills against measurable performance indicators. Indirect assessments are those that ascertain the opinion or self-report of the extent or value of learning experiences”. Direct assessments are either norm-referenced or criterion-referenced. Norm-referenced assessments are based on a set of assumptions that permit comparison of one individual’s performance to others who have completed the same assessment. This allows interpretations of scores relative to the performance of others (e.g., “this student has performed above the average”). Norm-referenced assessments generally consist of dichotomous items – those with one clear, correct answer, such as the selected-response questions that are common on tests, quizzes, and examinations. Generally, in a norm-referenced test the individual test-taker earns a certain number of points for each correct answer, and the scorer totals the number of points earned for the correct answers to create a score. The assumption is that the individual’s score represents the individual’s knowledge of the subject matter being tested. So, the higher the score, the more knowledge the individual possesses. Based on this assumption, scores can be compared among individuals. Criterion-referenced assessments are very different. They are designed to compare a student’s performance to a particular standard or criterion. This allows interpretations of scores in relation to the body of knowledge. (e.g., “this student has met the specified performance standard”.) Test takers are given a task, and their response - performance, behavior, or a final product – is assessed for the degree to which it meets certain levels of quality. Measurement of these types of assessments is done largely through expert judgment by individuals qualified to review the response, usually a teacher, professor, or

other disciplinary expert. The resulting measurements are not intended to be used to compare achievement among those who complete the assessment, but rather the degree to which an individual meets the criteria established for the task. These assessments are often measured using rubrics.

Self-assessment is the process whereby learners simultaneously create and undergo the evaluation procedure (Stanchina&Holec, 1985). Self-assessment is often recognized by Hancock (1994) as: ‘an ongoing process involving the students and teacher in making judgments about students ‘progress in language using non-conventional strategies’ (p.3). According to Brown and Harris (1997) self-assessment is an effective means of developing independent learning and helpful in changing perceptions of learning (as cited in Lee &Coniam, 2013). Underhill (1987) explains two kind of scales used for introspective self-assessment as *defined* versus *non-defined* and *general* versus *specific*. A *non-defined* scale has descriptions only at the top and bottom ends of the scale, whereas a *defined* scale has descriptions for every level in between. A *specific* scale asks the learner to rate his performance in a particular language situation, while a *general* scale refers to the use of language in general terms only. There are also practicality issues regarding using self-assessment: self-assessment can save the teacher correction time and support students in dealing with their progress. Investigations concerning self-assessment in language learning in general have examined the value of self-assessment in proficiency testing participants of all ages (Blanche & Merino, 1989). Paris and Paris (2001) believed that self-assessment was an effective strategy to help learners perform self-regulated learning because learners were better able to evaluate their learning conditions if they could assess themselves during the learning process. They can also further monitor and correct their course of learning, and as a result, improve their learning effectiveness. Self-assessment is also the major feature of the assessment-centered learning environment (Bransford, Brown, & Cocking, 2000).McConnel (2000) mentions the self-assessment can result in major benefits both for teacher and students, and it is more oriented to students, reduces some of the teacher’s load,provides instant feedback and helps to remove certain ‘barriers’ between teachers and students. Furthermore, the students become less dependent on their teachers, responsible and autonomous; they take on a more proactive role and develop self-confidence, while the teachers can evaluate the effects of their teaching efforts more accurately and more objectively. Responsibility can be additionally promoted by permitting students to agree the rules for self-assessment among themselves or together with the teacher. In this way, students are actively involved in decision-making process about evaluation criteria and the evaluation process of their own and

other students' works. Involvement of students in their own assessment is an important part of preparation for life and work setting.

Teacher assessment, on the other hand, can be carried out either on the spot or as a continuous assessment over a period of time. For a continuous assessment, the teacher's judgment is formed as a gradual process rather than as a sudden decision. Each time the learner attempts a task in class, the teacher has, in effect, administered a single constitutes a complete test of proficiency, and unlike a short oral test, will not be influenced by short-term individual variations such as nervousness, illness or fatigue on a particular day (Underhill, 1987, p. 27). Flemming and Stevens (2004) state that "for a teacher to use assessment effectively it is necessary to be thoroughly 'beside the wheel' and involved in the process of learning. It is easy to get so submerged in the language of assessment (summative/formative; normative/criterion referenced; validity/reliability) that one loses any real insight into the tensions and issues at the heart of the process."

Evaluation:

Evaluation is another term, which will be discussed in this chapter. Trochim (2006) defined evaluation as a systematic acquisition and assessment of information to provide useful feedback about some object. Another definition is made by Iseni (2001) as following: "Evaluation is the wider process of interpreting data to make judgments about a particular program or programs. It is the process of collecting, analyzing and interpreting information about teaching and learning in order to make informed decisions that enhance student achievement and the success of educational programs." Evaluation uses both quantitative methods (tests) and qualitative methods (observations) for gathering information (Republic of Tunisia Ministry of Education).

Iseni (2001) states that evaluation has two main purposes which are 'for learning and development.' Fredericks (2005) states that "effective evaluation is a continuous, on-going process." Also he adds, "Evaluation must be a *collaborative* activity between teachers and students." As regards students he says that "they must be able to assume an active role in evaluation so they can begin to develop individual responsibilities for development and self-monitoring" by mentioning that "evaluation needs to be *authentic*".

Whether Assessment or Evaluation?

The terms assessment and evaluation are often used interchangeably, in actuality they are two parts of the same process. Assessment is the process of gathering evidence of what the child can do. Evaluation is the process that follows this collection of data, including analysis and reflection, as well as decisions based on the data (Foundation for the Atlantic Canada English Language Arts Curriculum). What is assessed and evaluated, how it is assessed and evaluated, how results are communicated, and how results send clear messages to students and others about what is really valued—what is worth learning, how it should be learned, what elements of quality are most important, and how well students are expected to perform (Foundation for the Atlantic Canada English Language Arts Curriculum).

Purpose of assessment and evaluation is to inform teaching and to promote and encourage learning—to promote optimal individual growth. In order to provide information vital to the teachers, assessment and evaluation must be an ongoing and integral part of the teaching/learning process. It is one continuous cycle consisting of collecting data, interpreting data, reporting information, and making application to teaching (Foundation for the Atlantic Canada English Language Arts Curriculum).

Different Perspectives in Preferences of Testing:

EFL teachers had strong beliefs of multiple assessments. Most of the respondents believed that multiple assessments were more practical than the traditional paper-and-pencil tests. All of them believed that EFL teachers will find it easier to assess students' learning achievement in listening, speaking, reading, and writing. They believed that teachers can easily diagnose students' difficulties in learning. All of them believed that students will effectively develop strategies on using English if multiple assessments are used (Chan, 2007). Standardized test play an important role with EFL curriculum evaluation as well as student evaluation. According to Brown and Abeywickrama (2010), a good standardized test is a product of empirical research and development that goes simply agreeing on particular standards or benchmarks. This type of test also includes systemic procedures for administration and scoring.

In the past decades, school entrance exams which comprised reading and writing exerted a strong influence on teaching and learning in Taiwan, resulting in classroom instruction which focus only on these two skills. Consequently, without instruction aimed at developing speaking and listening skills, most students in Taiwan study English for years, but are not able to learn spoken English or express themselves fluently in English (Chen, 2003).

Such classroom focus has resulted in a “teach to test” approach, preparing students for assessment rather than developing their skills within well-planned a curriculum based on student and institutional needs or demands (Chan, 2007).

MEASUREMENT

Measurement experts may have improved their understanding about the testing practices needed to achieve valid results, but a variety of factors place pressure on the educators who decide which students to include on various tests (Hakuta, 2000). Furthermore, deciding on which measurement tool should be used is another issue because there are different types of measurement tools.

A rubric is a measurement tool that describes the criteria against which a performance, behavior, or product is compared and measures. Rubrics list the criteria established for a particular task and the levels of achievement associated with each criterion. These are often developed in the form of a matrix (Brophy, 2012). Rubrics are composed of four basic parts (Hawaii, 2012). In its simplest form, the rubric includes *a task description* (outcome being assessed or instructions students received for an assignment), *the characteristics to be rated*, *rows* (the skills, knowledge, and/or behavior to be demonstrated), *levels of mastery/scale*, *columns* (labels used to describe the levels of mastery should be tactful but clear), and *the description of each characteristic* at each level of mastery/scale.

There are two types of rubric:

Analytic Rubric: An analytic rubric presents a description of each level of achievement for each criterion, and provides a separate score for each criterion. It provides more detailed feedback on student performance and more consistent scores across students and raters. However, it may be more time consuming than applying a holistic rubric. Therefore, it can be used when you want to see strengths and weaknesses and detailed feedback about student performance.

Holistic Rubric: A holistic rubric presents a description of each level of achievement and provides a single score based on an overall impression of a student's performance on a task (Carriveau, 2010). It is quick to score, and it may provide an overview of student achievement. Moreover, it may be efficient for large group scoring. On the other hand, it may

not provide detailed information. Furthermore, it may be difficult for scorers to decide on one overall score.

Conclusion:

Over the decades, we have attempted to motivate by holding schools accountable for scores on standardized tests and by intensifying the stakes associated with low test scores. This began in the 1940s with college admissions tests. Next came district-wide standardized tests in the 1950s and 60s. The 1970s was the decade of the state assessment. In the 1980s and 1990s, we added national and international assessments (Chappuis, 2004).

Contrary to students' beliefs, tests are not just a means to complicate students' lives. The grading or assessment of tests can be used to enhance learning and teaching. Students spend more time studying outside of class than they would spend in the class-room for the exams. Therefore, tests and assignments are not only evaluation methods, but also tools that increase student learning (Wolfe, 2004). Not only is there little evidence that these multiple layers of externally-imposed tests have improved school quality or reduced achievement score gaps, some contend that they have exacerbated the problem by forcing increases in dropout rates and declines in graduation rates, especially among minorities. There are factors that can cause a student's score on a test to misrepresent his or her real achievement. Problems can arise from the test, the student, or the environment where the test is administered (Chappuis, 2004). As seen test taker's achievement can be affected by various factors.

Besides that there are many other factors which may lead to inaccurate test results. Tests can consist of poorly worded questions, place reading or writing demands on respondents that are confounded with mastery of the material being tested, have more than one correct response, be incorrectly scored, or contain racial or ethnic bias. The student can experience extreme evaluation anxiety or interpret test items differently from the author's intent, as well as cheat, guess, or lack motivation. Any of these could give rise to inaccurate test results. Or the assessment environment could be uncomfortable, poorly lighted, noisy, or otherwise distracting (Chappuis 2004). In settings where there is a statutory end-of-course exam to be passed, initial self-corrected tests can be useful. Tests can be administered which are similar to the end-of-course exams, e.g. by using past papers. Students can then be given

keys to work out their own scores. This can help them see what will be expected of them and how they have performed as individuals. They can also be asked to do written or oral tasks, and to compare what they have done with examples of successful and unsuccessful performance in previous exams (Harris, 1997).

References

- Bachman, L. F. (2011). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996) *Language Testing in Practice*. Oxford: Oxford University Press.
- Blanche, P., & Merino, B. J. (1989). *Self-assessment of foreign-language skills: implications for teachers and researchers*. Language Learning.
- Bransford, J.D., Brown, A., & Cocking, R. (2000). *How people learn: mind, brain, experience and school*. Washington DC: National Academy 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.
- Chan, Y.C. (2007). Elementary school of EFL teachers' beliefs and practices of multiple assessments. *Reflections on English Language Teaching* 7/1, 53.
- Chappuis, J. & Stiggins R. (2004). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice* 44, 5.
- Crocker, L., & Algina J. (2006). *Introduction to classical and modern test theory*. Thomson Learning.
- Ćukušić, M., Garaća, Ž., & Jadrić, M. (2014). Online self-assessment and students' success in higher education institutions. *Elsevier Journal*, 72.
- Flemming, M. & Stevens, D. (2004). *English Teaching in the Secondary School*. (2nd ed.) (pp. 113-114) London: David Fulton Publishers Ltd.
- Foundation for the Atlantic Canada English Language Arts Curriculum: K-3. *Assessing and Evaluating Student Learning*.
- Fowler, F. J. Jr., (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Hakuta, K. (2000). Testing. *Testing English Language Learners Book*.

- Hancock, C. (1994). Alternative assessment and second language study: what and why? Eric Digest. Retrived from <http://www.ericdigest.org/1995-2/language.htm>
- Hanna, G.S., &Dettmer, P.A.(2004).*Assessment for effective teaching: Using context-adaptive planning*.Boston, MA : Pearson A&B.
- Harris, M. (1997).Self-assessment of language learning in formal settings.*ELT Journal, Oxford University Press* 51/1, 14.
- Hubackova, S., Semradova, I., &Klimova, B.F. (2011). *Blended learning in a foreign language teaching* .Elsevier Journal, 28.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Iseni, A. (2011) Assessment, Testing and Correcting Students' Errors and Mistakes. *Language Testing in Asia* (1/3).
- Ivanova, Vanya (n.d). Construction and Evaluation of achievement tests in English.
- Khonbi, Z. A., &Sadeghi, K. (2013). The effect of assessment type (self vs. peer) on Iranian university EFLstudents'course achievement. *Procedia – Social and Behavioral Sciences*, 70.
- Learning Common Fastfact Series (2005). *Multiple Choice Exams, from www.learningcommons.uoguelph.ca*
- Lee, I. &Coniam, D. (2013).Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong.*Journal of Second Language Writing*, 22, (pp. 34-50).
- McConnell,D. (2002). TheExperience of CollaborativeAssessment in e-Learning. *Studies in ContinuingEducation*, Vol. 24 (1).
- Medwell, J., Wray, D, Minns, H., Coates E. & Griffiths (2009) Primary English (4th ed.)(pp. 168-169). Wiltshire: Penguin Books Ltd.
- Miller, M.J. Reliability and validity.*Graduate Research Methods, Western International University*.
- Ozerova, A. (2004), "Types of Tests Used in English Language. Bachelor Paper", S.(2, 4, 8, 9, 10, 11, 12, 13, 14, 15, 16)
- Paris, S. G., & Paris, A. H. (2001).*Classroom applications of research on self-regulated learning*.EducationalPsychocologist.
- Rahimi, M, ,Momeni, G., &Nejati, R. (2012). The impact of lexically based language teaching on students' achievement in learning English as a foreign language. *Elsevier Journal*,31.
- Reimann, A. (2003) *Role Play: Viable, Communicative Language Testing*. JALT 2003 at Shizuoka Conference Proceedings.
- Republic of Tunisia Ministry of Education.*Monastir Regional Board of Education Academic Year 2001-2002*

- Simith, S., Avinesh, P.V.S., & KilgarriFF, A. (2010). Gap-fill Tests for Language Learners: Corpus-Driven Item Generation. *Proceedings of ICON-2010:8th International Conference on Natural Language Processing*, Macmillan Publishers, India. [Http://ltrc.iiit.ac.ac.in/proceedings/ICON-2010](http://ltrc.iiit.ac.ac.in/proceedings/ICON-2010).
- Tierney, R.D.(2014). Fairness as a multifaceted quality in classroom assessment. *Elsevier Journal*.
- Tomlinson, B. (2005). Testing to learn: a personal view of language testing. *ELT Journal*, Oxford University Press 59/1, 44
- Underhill, N. (1987) *Testing Spoken Language: A handbook of oral testing techniques*. (1st ed.) (pp. 21-86).Cambridge University Press.
- Wolfe K. (2004). Student assessment-testing and grading. *Tips for Teaching Assistants and New Instructors. Journal of Teaching in Travel & Tourism*, 4/2, 80.