The Role of Pedagogical Testing and Assessment
In the Context of ELT.

Asst. Inst.Ra'ad Abdulhussein Hamoud Thaidan.

Ministry of Education
Directorate of Education in Dhi-Qar

The Role of Pedagogical Testing and Assessment In the Context of ELT.

Asst. Inst. Ra'ad Abdulhussein Hamoud Thaidan.

1- Abstract.

In this paper, the researcher discusses pedagogically the current role of testing and assessment scholastic domain which played a critical and an effective role in education and society. These pedagogical tools enabled the educational policy-makers to reform and rehabilitate the applications of testing in the educational settings.

Moreover, policy-makers resolve and recalibrate low school achievements for instance relying on the empirical studies regarding the power effect of examination. Educators and candidates are required to make progress with instruction following the scopes of testing and assessment agenda. Henceforth, in testing and assessment domains, test-takers are considered principally as important stakeholders whose views of assessment and testing applications are taken seriously into account in terms of setting criteria of testing a good test. In turn, these procedural steps onward enhance and facilitate assessment method and can influence students' cognitive performance moreover, to investigate such educational process forwardly. In accordance with this scholastic process, the assessment regimes and testing benefits can play a vital role regarding how to develop or explore candidates' potential skills as well as to reduce test-takers' attitudes or feeling of discomfort into more appraisal ones. Additionally, we can see how assessment influences the educational learning system, paving the way to take further decisions for developing the pedagogical purposes.

Key words: pedagogical test, test-takers, educators, curriculum objectives, policy-makers, and testers.

الخلاصة

في هذا البحث، يناقش الباحث من الناحية التعليمية الدور الاني لموضوعة القياس والتقويم المدرسي باعتباره المجال الذي يؤدي دور مؤثر وذو مغزى تربوي في ميدان التعليم والمجتمع. أن

هذه الادوات التعليمية كانت قد مكنت صناع القرار التربوي لغرض الخوض في مسألة إصلاح واعادة تأهيل التطبيقات ألاختبارية في مجال العمل ألتربوي والتعليمي.

فضلا عن ذألك، فأن صناع القرار في هذا المجال يعملون على اصلاح واعادة تقويم الأداء المدرسي لاسيما المنخفض منه على سبيل المثال معتمدين بذألك على ألدراسات التجريبية والمتعلقة بأثر وفعالية الامتحانات ألتقويمية. أن التربويون والاشخاص المنخرطين ضمن هذه العملية بحاجة للوقوف على سير وديمومة العملية التعليمية والتقويمية منها والتي تتبع مسألة القياس والتقويم قدماً.

يعد مؤدو الاختبارات الامتحانية في مجال القياس والتقويم أفرادا مهمين وذو علاقة أساسية مع هذا الجانب التحصيلي حيث أن وجهات نظرهم وآرائهم في ميدان تطبيقات القياس والتقويم تؤخذ على محمل الجد في مجالات معايير أعداد الاختبار التقويمي ذو المواصفات الجيدة.

في المقابل، فأن هذه الخطوات الاجرائية تعزز وتيسر طريقة التقويم قدماً ومن خلالها ممكن أن تؤثر على سير الإداء ألمعرفي للطلبة وكذألك يمكن أن تبحث في مثل هذه العملية التربوية ذات المغزى. بالتناغم مع هذه المسيرة المدرسية، فأن الأنظمة ألتقويمية والفوائد ألاختبارية يمكن أن تؤدي دور حيوي وفعال بما يتعلق بكيفية تطوير واكتشاف المهارات ألكامنة للطلبة وكذألك الأفراد ذوي العلاقة بالجانب التربوي هذا فضلاً عن تقليل موقف المتعلمين من ناحية مشاعر عدم الرضا الى جانب اكثر انتاجاً مثل الثناء والرضا عنها.

فضلاً عن ما اثير سالفأ، نستطيع أن نلاحظ كيف أن التقويم التربوي يؤثر على الجانب التربوي والتعليمي تباعاً، أذ يمهد الطريق لغرض اتخاذ المزيد من القرارات بعيدة الاثر لأغراض تطويرية.

2- Introduction.

It is knowledgeable that, the regime of assessment and testing domains has principally a bifunctional nature which determines the negative back draws or positive yields regarding to teaching and learning processes. The interest in classroom assessment and setting scholastic tests come to the prominence in order to daunt the pitfalls of instructional and assessing agendas, essentially.

However, the great contributions of applied linguistics enhance the field of English language teaching in terms of discriminating the traditional or classical mode of language measurement to provide a basic pedagogical foundation psychometrically as well as to further candidates' learning process, powerfully.

Debatably, the most effective challenge to language testing agenda is the concern of the context in which testing process is carried out bifunctionally at the micro and macro level, accordingly. Henceforth, the micro domains stress critically on the recent thinking agenda of classroom assessment. In turn, the macro scope of social context awareness has contributed the developmental strategy of critical language testing as well as testing construct.

Therefore, to drive scholastic testing and assessing process timely, an alignment and conformity of testing strategies played and echoed a vital important consequential impact of performance assessment. Then, the themes of this paper are to shed pedagogical lights on how to create knowledge concerning effective assessment for learning moreover, assessment as ongoing learning process.

It is obvious to know that, subsequent related researches targeted the back draws competence of in service or under service educators in the assessment areas for improvement purposes and rehabilitate them. Consequently, realities of classroom assessment come to the prominence of education relevantly to understand and interpret how to design, construct, and understand the cognitive elements that related to assessing and testing students' ongoing learning.

3- The Purpose of Pedagogic test and Assessment.

The core of educational test and assessment is considered as an intermingled pedagogical phenomenon. This ongoing educational process includes variable pedagogical functions in the educational role regarding to the potential entity of society. Today, language testing use provides a systematic educational scopes and a wide speared educational strategy in terms of pedagogical policy making.

In accordance with this trend, Fox and et al (2007:145) state that, "the currently reality of tests, not only of language tests, is that they are

given more power than ever, as they are widely used by governments, institutions, and central authorities world wide". However, as language tools and a powerful domain, language testing played a critical role in the educational domains for perpetuating variable agendas, dramatically.

It is knowledgeable that, English language tutors employed tests to measure and diagnose learners' background schematic knowledge for instance as well as they assess formatively or to use summative mode of learners' progress regarding to curriculum setting or instruction concerns. In this sense, Mc Millan (2013:56) agrees that, "the results from interim assessments are frequently used to provide diagnostic feedback to aid instruction as well as provide predictive information regarding how a student is likely to perform on year-end assessments".

On the other hand, the pedagogical agendas of language tests as unavoidable practice played a significant and powerful role in terms of implementing the pedagogical opportunities to learner's progress. Bachman and Palmer (2000: 177-178) conclude that, "we cannot be sure that multiple forms of a test will provide equivalent measures of test-takers' abilities until we have tried them out and analyzed their results". In this sense, the educators in the educational domains practise the pedagogical assessment as a traditional ongoing process in order to gauge or to gather information concerning learners' background knowledge.

At the same time, as educators, we can judge about the instructional effectiveness and the curriculum efficiency. Here, Celce and et al (2014:307) maintain that, "increasingly, however, information from large-scale assessments is being used for diagnostic purposes-namely, to assess test-takers' strengths and weaknesses so that future teaching (or learning) can be prescribed". Henceforth, language teachers need urgently to gather information and make their educational standard judgments on the candidates' progress in terms of students' achievement within a particular context so as to diagnose classroom difficulties and expected future obstacles regarding to input or output concerns.

According to the previous scopes, Caroll (1961: 314) puts it that, "the purpose of language testing is always to render information to aid in making intelligent decisions about possible courses of action. But, these

decisions are diverse, and need to be made very specific for each intended use of a test" (cited in Fulcher, 2012:1). So, the procedural process of a test and assessment enhances the educational domain in terms of making inferences concerning specific characteristics of students for instance. Berry and Adamson (2011:3) report that, "assessment are required to provide important information for the purposes of selection; for government ministers, assessment results enable them to evaluate the effectiveness of the education system and how it compares with that of other states through the rankings in comparative studies".

It is obvious that, classroom tests played a considerable educational role in learning and teaching domain. Such process of testing and assessment are employed by educators and candidates regarding to which extent the curriculum objects have been achieved for example. Berry (2008:6) agrees that, "assessment is a deliberate and planned collection of the full range of information from the students that helps them understanding their knowledge, skills, and abilities, including strengths and weakness, values, and attitudes, assessment is a natural part of the teaching and learning process and is undertaken to support learning" . Henceforth, the process of tests and assessments play critical and powerful tools regarding to privilege certain levels and forms of language knowledge. Thus, Paran and Sercu (2010:1) conclude that, "language tests, give their power and influences in societies; play a major role in the implementation and introduction of language policies".

Therefore, English language teachers can imply the term 'assessment' in the educational scholastic setting as a method for better understanding of their candidates' current linguistic knowledge. In this commentary, Cowie, Moreland and Otrei-Cass (2013:18) agree that, "the social and public nature of classroom assessment is highlighted and linked to its power in bestowing, sustaining and transforming student identities as learners in the classroom setting".
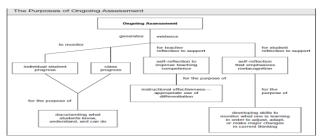
Figure (1) Ongoing Assessment (Tomlinson and Moon 2013:77).

It is knowledgably that, ongoing assessments allow us to gather a pedagogical judge regarding to learners' performances or progresses. To promote learning, ongoing assessment requires to set a clear learning goals, a specific information regarding to the learners relative to those goals and take an effective steps further to equip teachers with sufficient skills and knowledge concerning their leaner's needs. In accordance with this, Douglas (2014:2) concludes that, "tests also allow other stakeholders, including programme administrators, parents, admission officers and prospective employers, to be assured that learners are progressing according to some generally accepted standard to have achieved a requisite level of competence in their second language".

4-The Place of Tests and Assessment in Education.

It is evident that, in the last three decades, a distinctive growing interest of procedural applications of testing and assessing in variable pedagogical domains, radically rather than classically. Authentic and promising forms of tests and assessment have come increasingly to the prominence in the ELT world. The norms of dealing with accumulated knowledge moves demonstrably from traditional mode of quizzes to standardized authentic tests for instance.

However, English language teachers and learners unfortunately in the educational settings opine tests accordingly as an unwelcome nightmare, upsetting them and decreasing their affective filters down due to the questions that they cannot answer or in terms of being unable to get the good grades. So, the experience of testing and assessment process can be negative on students' achievement and can be positive in terms of building students' potential confidence forwardly. Therefore, Brown and Hudson (2003:271) believe that, "those who are writing the test will need to determine if the first goal is primarily an enabling skill, that is, a

mechanism for getting to the second goal or an important end goal that will be necessary for the next level of language courses".

Arguably, Mc Namara (2000:69) puts it that, "assessment serves policy functions in educational contexts, too. One example is in the area of vocational education and training for adults". Henceforth, Birjandi and et al (2004:109) maintain that, test outcomes provide testers and planners with variable information concerning the efficacy of educational programming objectives as well as the positive and negative backwash points of the course procedural steps inwardly or backwardly.

However, since 1980s, there has been a distinctive engagement in the important role of assessment which takes place, distinctively in the process of synthesizing classroom information rather than at the end of a course sessions. In accordance with this, Elder and et al (1998:110) conclude that, "we thus believe that test usefulness provides not only a metric by which we can evaluate the tests we use, but also on essential basis for quality control throughout the entire test development process".

It is understood that, as a pedagogical instrument, administrators employ 'a test' to elicit candidates' performance to discover to which extent the learners have attained the course objectives. Brown (2000:401-402) thinks that, "tests are almost always identifiable time periods in a curriculum when learners muster all their faculties to offer peak performance, knowing that their responses are being measured and evaluated".

But, on the other hand, assessment encompasses a much wider field as compared with testing domain. English language teachers depend on testing or assessing process to improve and raise standards of instructional settings. This would lead policy-makers to align learning and teaching activities into better agenda in order to meet the pedagogical need of students. Bachman and Palmer (2000:10) maintain that, "we need a conceptual framework that enables us to treat performance on a language test as a particular instance of language use. That is, we need a framework that enables us to use the same characteristics to describe what we believe are the critical features of both language test performance and non-test language use".
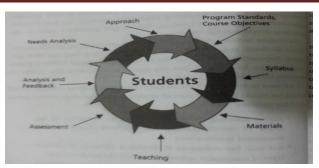
Figure 2: Assessment in teaching / learning process (cited in Coombe 2010:.2)

Figure (2) explains the critical role of assessment as an integral part of the potential cycle of curriculum. Therefore, the pedagogical decisions concerning how to assess should be targeted from the early beginning of curriculum design or syllabus planning. Here, Jo Mc Donough and et al (2013:269) state that, "we might also suggest that learners do not learn directly from a syllabus, but what they learn, or not, is the result of the manner in which this syllabus is 'translated' into the classroom environment, in the form of materials but also of their use by the teacher and learners in the class".

It is understood that, both of testing and assessing process can enhance and develop forwardly the instructional domains. Pedagogically, assessment can be helpful to feedback and analyze the components of syllabus objectives. Coombe and et al (2010:3) add that, "analysis is the basis for helpful feedback to students, teachers, and administrators. Assessment coupled with analysis can improve instruction; assessment alone cannot".

Then, Wyatt-Smith and et al (2014:84) agree that, "teachers need to learn how to re-organize their class to enable ability grouping for instructional purposes and how to use data to make instructional decisions".

5-Washback Effect in Testing Settings.

It is suggested that, the outcomes of any given tests can be used for variable purposes like diagnosing the different domains of the educational ongoing process or to aid learning for instance. Language testers are engaged in this dramatic access in terms of rejecting or supporting the functional use of tests in the educational settings. Tests can be helpful determining the extent to which the curriculum or instructional objectives

were realistic as well as whether the contents and methods of instruction were organized, appropriately or not. However, the pedagogical outcomes of testing and assessing process reveal the instructional weakness of test-takers and also the negative practices of instruction accordingly.

Regarding to this, Fox and et al (2007:150-151) state that, "we need to get engaged in the potential debate about language awareness and language testing activism so we can influence those without better access to power centers, educational systems, law making, and especially the implementation of language-related policies".

Henceforth, the test would be regarded as an important and an effective tool if the test stakes are considerably high. On the other hand, if the contents and techniques of testing derail and mismatch the predetermined objectives of the course, there will be harmful washback. So, it is important to identify about the changes in terms of teaching and learning which can be attributed directly or indirectly regarding to the use of the test in the educational contextual practice.

In accordance with this, Hughes (2003:2) voices that, "the proper relationship between teaching and testing is surly that of partnership. It is true that there may be occasion when the teaching programme is potentially good and appropriate but the testing is not, we are then likely to suffer from harmful backwash".

It is interesting to note that, the testing scholastic domain plays an influential effect in the educational systematic settings. The classroom examinations had a distinctive power on the stakeholders in terms of negative or positive impact. Loumbourdi (2014:10) stresses that, "the power and authority of tests enable policy-makers to use them as effective tools for controlling educational systems and prescribing the behavior of those are affected by their results-administrators, teachers and students".

It is noted that, when most of candidates do poorly on the same exam items, this may not correlated with testees' educational faults, but the drawback is possibly reincarnated into the efficiency of instruction or applying inappropriate teaching method. Mc Namara (2000:74) puts it that, "as performance assessments required integration of knowledge and skills in performance on realistic tasks, preparation for such assessments will

presumably encourage teachers and students to spend time engaged in performance of such tasks as part of teaching".

Therefore, it is worth of note that, there is an urgent need to raise a critical awareness regarding to potential testing washback effects in order to seek and capture the social dynamics properties between the entity of tests and the testees. Finally, Bachman and Palmer(2000:352) conclude that, "the validity of test use is based on reliability. The tests which do not have these qualities will not be able to insure the appropriateness of interpretations of test us for the intended purposes". Shohamy (1993: 37) furthers and argues that, "testers must begin to examine the consequences of the tests they develop…often…they do not find it necessary to observe the actual use of the test" (cited in Weir, 2005:38).

6-The Pedagogical Scopes of Language Testing.

It is interesting to note here that, today, testing and education are still playing a critical philosophical roles and a significant impact in the pedagogical political policies. Testing as an effective classroom tool has been employed particularly as an educational systematic pile for policy decision makers in order to steer the educational system, significantly and to make appropriate pedagogical adjustment to meet learning needs.

Hence, the primary scope regarding to develop a language test is the use and utility for which it is intended. The qualities, the usefulness of a test and the test settings are required components to policy makers in order to make their inferences, predictions, and decisions, pedagogically. In accordance with this view, Elder and et al (1998:109) explain that, "we usually want our inferences about language ability to generalize beyond the test itself, so that we need to carefully consider that extent to which the performance we elicit in a language test corresponds to the language that is used in non-test language use settings".

Bachman and Palmer (2000:17) agree with this trend believing that the usefulness of a test provides a type of metric by which we can evaluate not only the test being developed and used, but all aspects of test development and use. Furthermore, Fox and et al (2007:152) argue that, "tests need to be used not just to penalize *bad and impure language* but to encourage the complex language varieties that are used among the diverse

populations in this world and by extension to avoid the imposition of unrealistic criteria that serve only the privileged".

It is obvious that, policy makers in the educational domains try and wish hardly to find a balanced pedagogical tests and assessment system like authentic or useable ones for the utility of candidates and teachers, accordingly. Developing language assessment is targeting the educational progress of learners. Richards and Renandya (2002:339) comment that, "although it is unlikely that any one instrument will fit the needs of a given group of students, the idea is to adopt and/ or adapt existing instruments in such a way that they reflect the goals of the class and the activities being implemented in that classroom to meet those goals".

Regarding to this view, Mc Millan (2013:8) maintains that, "given the current climate of school accountability and data-driven decisions making, teachers are expected to use sound assessment practices as a means for improving student outcomes. They need to integrate large-scale testing approaches and item types with scale testing approaches and item types with how they assess their students and plan for future instruction".

The implementation of language testing polices enables decision makers and central authorities to diagnose the back draws sides for remedial purposes. The outcomes of class test performance will enhance learners' progress regarding to interpret the instructional system. Here, Weir (2005:14) concludes that, "we can collect data on events after the test has been developed and administered to shed further light on the well-foundedness of the inferences we are making about underlying abilities on the basis of test results".

Therefore, there is an urgent need to design and align test tasks to be authentic and reliable. This claim comes to the prominence due to the philosophical trend among variable test qualities. This philosophy in English language learning and teaching shifts under such justifications. Bachman and Palmer (2000:18) maintain that, "we would thus argue that test developers need to find an appropriate balance among these qualities, and that this will vary from one testing situation to another".

Bachman (1995:56) concludes that, "the amount and types of testing, if any that is done depends upon the decisions that are to be made

and the type of information that is needed to make the correct decisions". According to the above mentioned discussion, English language testers or assessors are progressively in need to understand the critical and significant role that language testing powerful domain play in creating the educational language policies in terms of ramifications, pedagogically in the basic educational skills to take an active role in the community.

7-The Criteria of Good Tests in ELT.

Knowledgeably, H. D. Brown (2004) identifies five cardinal principles or factors regarding to 'testing a good test' in English language teaching, namely: Practicality (usability), Reliability, Validity, Authenticity, Washback (or Backwash).

7-1 Practicality (Usability).

An effective test should be practical. This feature reincarnates into variable means of feasibility in administration, scoring and interpretation of its expected outcomes. Moreover, it should not be excessively expensive. D. H. Brown (2000:386) states that, "a test that is prohibitively expensive is impractical".

So, usability can be defined as "the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these activities (Bachman and Palmer, 2012:262). A test may include features of reliability or validity, but it considers as impractical one when it consumes more time and money than necessary to achieve the predetermined educational objectives. Henceforth, Birjandi and Mosallanejad (2010:190) comment that, "a test that needs very expensive equipment, or unique situations to be administered, does not have the ease of administration". So, subjective mode of testing for instance as compared with objective ones is difficult to score. Here, two scorers may provide variable scores. This happens due to the difficult opinions regarding to test results interpretations, accordingly.

However, the extent to which a test is usable swings accordingly in terms of test design into two distinctive modes. The first one is *norm-referenced* while the other is called *criterion-referenced.* Regarding to norm-referenced test, the candidates' obtained scores are interpreted in

relation to a mean, medium, standard, and percentile level in order to place test-takers in rank order.

H. D. Brown (2000:386) furthers that, "typically of norm-referenced tests are standardized tests intended to be administered to large audiences, with results quickly disseminated to test-takers. Such trend of tests must have fixed predetermined responses in a formal that can be electronically scanned". In return, criterion-referenced tests are designed to give testees backchannel (as a form of feedback) on specific scholastic course or lesson objectives for instance, that is, the 'criteria'. H. D. Brown (2000:386) explains that, "classroom tests involving smaller numbers, and connected to a curriculum, are typically of criterion-referenced testing". In accordance with this, to deliver the feedback concerns, more time and efforts are required to be applied by testers (teachers) accordingly.

7-2 Reliability.

It is knowledgeable that, this criterion is deemed as a critical feature of any good test. According to Croker and Algina (1989:105) "whenever a test is administered, the test user would like some assurance that the results could be replicated if the same individuals were tested again under similar circumstances" (cited in Fulcher and Davidson 2007:104). Any given test for example cannot measure anything without measuring it consistently (or reproducibility).

Bachman and Palmer (2000:19) visualize this principle as a consistency of measurement regarding to test score across variable features of testing conditions. So, consistency is seen as a critical and logical issue. However, factors or roots of unreliability may brim in the procedural test process itself (known as test reliability) or in the scoring affairs (known as rater reliability). Imagine that, fifty examinees take the same test with the same subject that is not difficult on different identical occasions, if the outcome of the test yields similar results, and then it should have test reliability. But, if the results yields difference factors, this may lead to contribute to unreliability of such given test.

Birjandi and Mosallanejad (2010:167) add that, "although all reliability coefficient are obtained by correlating two sets of scores obtained from the same test or from equivalent forms of the same test, the

results will vary with the specific method used". Consequently, Genesee and Upshur suggest a table, categorizing the kinds of reliability and way of enhancing it.

Type of reliability	Ways of enhancing reliability.
-Rater reliability	-Use experienced trained raters. -Use more than one rater. -Raters should carry out their assessments independently.
-Person-related reliability.	-Assess on several occasions. -Assess when person is prepared and best able to perform well. -Ensure that person understands what is expected (that is, those instructions are clear).
-Instrument-related reliability.	-Use different methods of assessment. -Use optimal assessment conditions, free from extraneous distractions. -Keep assessment conditions constant.

Table 1: Types of reliability (Genesee and Upshur, 1996:60).

Consequently to the above mentioned details and according to the views made by Genesee and Upshur, examinees need to be provided with accurate rubrics with clear descriptors regarding to what they have to know in terms of obtaining grades. Such sufficient instructions will fairly make them familiar enough with their educational upgrades in order to meet the standardized aspects of testing domains prior to engaging in the process, forwardly. J. D. Brown (2005:220) concludes that, "reliability can be viewed as a precondition for validity that is a test cannot be valid unless it is first reliable".

7-3 Validity.

This important criterion in testing domain occupies a vital and important pile of test component. Mousavi (2009:815) defines this term stating that, "a test is said to be valid to the extent that it measures *what it is supposed to measure* or can be used for the purposes for which it is intended". In the world of testing and assessment, Fulcher and Davidson

(2007:4) explain that, this principle has been classically used to reveal and discover whether a test measures what is supposed to be measured, accurately or not.

A test of reading comprehension for instance, should test reading comprehension knowledge and capacity of learners in reading related issues. J. D. Brown (2005:220) explains that, "a test can be reliable without being valid. In other words, a test can consistently measure something other than that for which it was designed".

However, to ensure validity factor in a test, we have to consider such case as a challenging task since the task end outcome should be taken as the actual information which mirrors the required characteristics that we intend to know about. Here, Shin and Crandall (2014:.252) argue that, "assessment have social and educational consequences, they must be fair, and what they are used for must match their intended purposes to be valid assessment".

As a result, Heaten (1990:159) explains that, a test must to provide a true measure of the specific skill which it is aimed to measure, the external extent potential knowledge as well as other skills accordingly, and then this will be seen as a valid test. In accordance with this, H. D. Brown (2004) categories and lists four cardinal aspects of validity which is complementing each other accordingly. These are: *content validity, construct validity, criterion-related validity, and face validity.*

7-3-1 Content Validity

Content validity is deemed as an important component of validity itself in terms of test achievement. H. D. Brown (2000:388) puts it that, "if a test actually samples the subject matter about which conclusions are to be drawn, if it requires the test-takers to perform the behavior that is being measured, it can claim content validity".

So, the standardized mode of test achievement may include high content validity because it resembles the common objectives in an area for example. But, in turns, it still includes low content validity in a specific situation. Birjandi and Mosallanjad (2010:184-185) maintain that, "to determine whether a standardized achievement test is valid for use in a

particular classroom, it is necessary to evaluate the items in high of the content and objectives which are emphasized in the instruction".

7-3-2 Construct Validity.

A second category of validity component that educators must be aware of regarding to language tests is construct validity. It is concerned, "the extent to which performance on tests is consisted with predictions that we make on the basis of a theory of abilities, or constructs" (Bachman 1995: 254-255).

Therefore, educators are interested in this category of validity when they intend to use test performance in order to gauge or infer the possession of certain qualities or psychological traits, forwardly. Birjandi and Mosallanejad (2010:186) provide a suitable example maintaining that, "instead of talking about a person's scores on test x, we want to talk about the person's intelligence, reasoning ability, or mechanical aptitude".

Consequently, these hypothetical features are termed constructs which are supposed to exist so as to take into account of behavior in many variable situations, pedagogically.

7-3-3 Criterion-Related Validity.

The third form of validity is also known as criterion-related evidence. H. D. Brown (2004:24) defines this mode of evidence as, "the extent to which the 'criterion' of the test has actually been reached". Educators employ this category of evidence so as to predict future performance as well as to determine current performance regarding to some other measure accordingly.

Here, Hughes mentions two types of criterion-related evidence namely: *concurrent validity and predictive validity*. According to Davies (1990:24), concurrent evidence is basically piled on, "a measure that is already at hand, usually another test, and in its most pure form concurrent validity can be established only when the test under scrutiny represents either parallel version of the criteria test or a simplified version of it".

On the other hand, Hughes (2003:29) maintains that predictive validity "concerns the degree to which a test can predict candidates' future performance".

7-3-4 Face Validity.

This concept is related closely to content validity. Hughes (2003:33) explains that, "a test is said to have face validity if it looks as if it measures what is supposed to measure". So, does the test on face validity of it appear from test-takers' perspectives regarding to test what is designed to test? For example, a test that designed to measure writing ability, but which did not require the testees to write, this mode of test lacks face validity.

According to Birjandi and Mosallanejad (2010:189), face validity refers primarily to"

    (a) The representativeness of test content,

    (b) evidence of validity from the psychometric perspective,

    (c) The superficial characteristics of a test,

(d)   The amount of empirical validation data accumulated for a test".

8- Authenticity.

The testees are supposed to perform positively when they engage to test tasks which contain the logical related modes of real-life situations which implemented in the target language classroom activities. Educators are required to depend on using authentic tasks to gauge test-takers' language skills, accordingly.

H. D. Brown (2004:28) puts it that, authenticity may be present in the following ways:

- " the language in the test is as natural as possible,
- Items are contextualized rather than isolated,
- Topics are meaningful (relevant interesting) for the learner,
- Some thematic organization to items is provided, such as through a story line or episode,
- Tasks represented, or closely approximate, real-world tasks".

9- Washback (Backwash).

The last major criterion of language testing is preferably known as washback or backwash. This principle comes to be used preferably in English language teaching because of the contribution of Applied Linguistics. So, in large-scale assessment, this term is defined by H. D. Brown (2004:28) which, "refers to the effects the tests have on instruction

in terms of how students prepare for the test". This denotes to the pedagogical testing process on teaching or learning affairs.

However, such concept played an effective role in conducting tests and examinations. Wall (1997) made a clear distinction between, test impact and test washback regarding to the terms of the scopes of the effects, accordingly. Here, Wall explains that, the term 'impact' denotes to, "any of the effects that a test may have on individual's, policies or practices, within the classroom, the school, the educational system or society as a whole whereas washback (or backwash) is defined as the effects of tests on teaching and learning" (cited in Cheng et al, 2004:4).

Therefore, the implementation of this principle in the educational setting will yield a positive and negative outcome. In accordance with this, the mismatch between contextual material, format, examination, and teaching managements would lead consequently into derailing regarding to implement curriculum predetermined objectives, forwardly.

It is understood that, researchers in the assessment and testing movement domains like Wall (1997), who made a pedagogical distinction clearly between what is already known as Macro-backwash (which denotes to the impacts on individuals, educational practices, and policy makers as well as Micro-backwash (which refers to the effects on learners and educators inside the school. In return to this, washback has the power to create a positive and negative domain in order to further or to enhance learning ongoing process or to face the obstacle of educational movement, strategically.

10- Conclusion.

The educational settings have always been a domain which assessment and test employed effectively in order to further and extend the strategic high scopes of educational agendas. The importance of scholastic assessment and tests practices provides and paves the educational basis for enabling researchers in testing development to deal with authentic world of assessment.

It is known that, English language learners hate to deal with unfair tests just like their teachers. The applied linguistics and educational measurement have been promoted and strengthened to meet the

stakeholders' pedagogical needs. This in turns, enabled those who are engaged with the field of ELT to apply the real world illusions of contextualizing testing domains in order to satisfy educational, economic, and political tends, systematically.

Previously, tests in the institutional centers were conducted regarding to their applications aggressively, biasing the classical agenda subjectively rather than objectively. Then, as a repair movement which targets the traditional mode of testing, test-developers, and researchers in the field of ELT. Linguists researchers Enhance new strategic scopes like large-scale assessment or finding a good criterion to gauge a good test, entirely for enhancement pedagogical purposes.

Teachers after a long period of educational practices are enabled recently to gather information about their learners depending on the modernized testing tools, effectively. Standardization method and approaches come to the prominence as systemic mode to testing and assessing learners for accurate educational judgments. This in turn, helps policy-makers to take steps in terms of building their pedagogical scopes regarding to planning, allocating resources, and developing curriculum objectives.

Then, this in return helped to enhance testing and educational system which encourages educators to emphasize on contextualizing learners' skills in testing scopes. In recent years, educators come closer to understand the gap between good and bad tests. They understood how to apply the practical educational standards to their teaching and learning process.

Consequently, this process has led directly to the pedagogical practices of (teaching to the test) regarding to set the assessment and to set high instructional benchmarks for individuals' learning results. Finally, the researchers of ELT have stressed on the critical importance of testing a good test which reincarnated into the criteria of a good test. These technical categories reported hence forwardly the growing importance of high-scale testing, dramatically.

REFERENCES

Berry, R. & Adamson, B. (2011). Assessment Reform in Education. Springer Dordrech

 Heidelberg London New York.

Berry Rita. (2008). Assessment for Learning. Hong Kong University Press.

Birjandi, P., Mosallanejad, P. (2010). An Overview of Testing and Assessment. Asfahan:

 Sepahan Press.

Brown, J. Dean. (2005). Testing in Language Programs. New York: McGraw-Hill Board.

Brown, J. Dean & Hudson, T. (2003). Criterion-referenced Language Testing. Cambridge:

 Cambridge University Press.

Brown, H. Douglas. (2004). Language Assessment Principles and Classroom Practices.

 Longman: Pearson Education Press.

Bachman Lyle F. & Palmer Adrian S.(2000). Language Testing in Practice (3rd.ed.). Oxford:

 Oxford University Press.

Brown, H. D. (2000). Teaching by Principles (An integrative Approach to Language

 Pedagogy). Longman University Press.

Bachman Lyle F. & Palmer Adrian S.(2012). Language Assessment in Practice (1st.ed).

 Oxford: Oxford University Press.

Bachman Lyle F. (1995). Fundamental Consideration in Language Testing (3rd.ed.). Oxford:

 Oxford University Press.

Cheng, L. Yoshinori, W., & Andy, C. (2004). Washback in Language Testing. London:

 Lawrence Erlbaum Association Press.

Celce-Murcia, M., Brinton Donna M.,& Snow, M. Ann. (2014). Teaching English as a Second

 or Foreign Language (4th.ed.). Boston: Heinle Cengage Learning Press.

Cowie, B., Moreland, J. and Otrel-Cass, K. (2013). Expanding Notions of Assessment for

 Learning. Sense Publisher ,The Netherlands

Coombe, C., Folse, K., & Hubley, N. (2007). Assessing English Language Learners. U.S.A.:

      University of Michigan Press.

Douglas Dan. (2010). Understanding Language Testing. Routledge Press: London&New York

Davies, A. (1990). Principles of Language Testing (1st.ed). U.S.A.: Blackwell Press.

Dietal, R. J., Herman, J. L., and Knuth, R. A. (1991). Assessing and Reporting the classroom

      curriculum. Oxford University Press.

Elder, C., Brown, A., Grove, E., Hill, K., Iwashita, N., Lumley, T., McNamara, T., O'Loughlin,

      K. (1998). Studies in Language Testing. Cambridge University Press.

Elder, C., Brown A., Grove N., Hill K., Iwashita N., Lumley T., McNamara T. & O'Loughlin K.

      (1998). Studies in Language Testing. BRT: Cambridge University Press.

Fulcher, G. (2012). Practical Language Testing. UK: Hodder Education Press.

Fox, J., Wesche, M., Bayliss, D., Cheng, L., Turner, C., and Doe, C. (2007). Language Testing

      Reconsidered. Ottawa University Press.

Fulcher, G. & Davidson, F. (2007). Language Testing and Assessment (1st.ed.). London:

      Routledge Press.

Genesee F. & Upshur John A.(1996). Classroom-Based Evaluation in Second Language

      Education (3rd.ed). Cambridge: Cambridge University Press.

Hughes, A. (2003). Testing for Language Teachers (2nd.ed.). UK: Cambridge University Press.

Heaten, J. B. (1990). Writing English Language Tests (3rd.ed). UK: Longman Press.

Jo Mc Donough, Christopher Shaw, and Hitomi Masuhara. (2013). Materials and Methods in

      ELT. (3RD.Edition). Wiley-BlackWell Press. U.S.A.

Loumbourdi Lambrini. (2014). The power and Impact of Standardised Tests. Peter-Long

      Edition: Frankfurt.

Mousavi, S. Abbas. (2009). An Encyclopedic Dictionary of Language Testing (4th.ed.).

      Tehran: Rahman Press.

McNamara, T. (2000). Language Testing (2nd.ed.). UK: Oxford University Press.

McMillan, James, H. (2013). Research on Classroom Assessment. SAGE Press. U.S.A.

Paran, A. & Sercu, L. (2010). Testing the untestable in Language Education**.** Short Run Press.

Richards, J. & Renandya, W. (2002). Methodology in Language Teaching (1st.ed.). London:

       Cambridge University Press.

Shin, J. Kang & Crandall JoAnn (Jodi). (2014). Teaching Young Learners English. Boston:

       Heinle Press.

Tomlinson, C. A. & Moon, T. R. (2013). Assessment and Student Success in a Different

       Classroom. Ascd Press .U.S.A.

Wayatt-Smith, C., Klenowski, V., Colbert, P. (2014). Designing Assessment for Quality

       Learning. Springer Press. U.S.A.

Weir, C. J. (2005). Language Testing and Validation (1st.ed.). BRT: Palgrave Macmillan Press.