



Fundamental Issues in L2 Classroom Assessment Practices

Volume 8 Issue 2 Spring 2010

Posted On Wed, Nov 30 -0001 00:00:00

Authors: [Esmaeel Hamidi](#)

Introduction

The educational reforms since the second half of the last (20th) century and their striking influences on the theories and principles of teaching and learning have also brought about a movement in assessment paradigms from measuring the amount of learning to enhancing learning which focused on more contextualized, communicative, performance-based as well as authentic assessment. This revolution, owing its outbreak mostly to the social constructivist framework of mind, has proposed to move toward helping learners to make their own decisions in learning. The very aspect of the reform has soon influenced all other areas of language programs such as instruction, curriculum design and materials development, educational administrative affairs, etc. In fact, assessment has taken on a high profile (Gipps, 1994). Thus, educators in the field have inexorably redefined the objectives of assessment both in the *global* and *local* levels. Today, assessment aims to support (not affect) learning and teaching, provide information about (not for) students, teachers, and schools, serve as both student selection and certification based on criteria (not on norms), act as an accountability (not an impact) device, integrate curriculum and instruction to assure [articulation](#), etc. This shift from product-oriented approaches to process-oriented approaches to assessment has very soon placed a lot greater demands on learners, teachers, parents, teacher trainers and developers, administrators, curriculum/materials developers, communities, and in short on all those in the state, district, and school levels.

The purpose of this paper is to provide L2 teachers with a succinct, conceptual framework of implementing the assessment forms to facilitate the relevant practical issues in L2 classrooms while focusing the current theories and views of assessment issues in L2 classrooms. I begin with giving a good rinse-out to *assessment* by distinguishing it from testing and evaluation. Then, following a brief account of the pedagogical history of assessment with a major focus on its 'authentic' aspect as the current concern to the educators of the field, I will discuss four criteria regarding the quality of assessment. The paper will continue with assessment purpose in two *levels* and assessment design in six *steps*. Moreover, fourteen principles will be pointed out. The paper concludes by arguing that L2 classroom assessment practices call for developing a profound understanding of the conceptual frameworks of the relevant defining issues.

Terminological delineation

One of the ways of finding out what L2 classroom assessment really means is to study its relationship with other terminologies such as evaluation and testing. This has been an issue of concern to many scholars. The reason why such distinctions are made is not the fact that other terms carry more weight than assessment, unlike certain scholars such as Cummins (1980) and Bachman (1990) who, overstressing the power of tests, claimed that a teacher's rating of a student's language based on informal interactive social language use may not be a very good indicator of how well that student can use language to perform various 'cognitive/academic' language functions; rather, it is the subtlety and broad-basedness of assessment which receives more attention. It is much more demanding than others. It is by nature the basic word. The point of note here is that most people in the field view assessment and any of them as dichotomies and use 'versus' to differentiate them, while they are not opposites. They are essentially correlated. They follow different goals, however.

- Assessment and Evaluation

Many people in the applied linguistics make no difference, or at least no special distinction, between these two terms. However, there are indications that are enough to distinguish assessment from evaluation. *Assessment* is a term that refers to a thorough but constant appraisal, judgment and analysis of students' performance through meticulous collection of information. *Evaluation* is described as an overall but regular judgment and analysis of teaching, learning, as well as curriculum through systematic collection of data. In assessment, the focus is on specific points of language; but in evaluation, the emphasis is placed on overall aspects of language. Assessment calls for forming a process which occurs during the learning process, but evaluation accentuates the conclusion of a process that takes place at the end of the term. Assessment looks at the individual language learners, but evaluation checks the whole language learning program. Assessment aims to inform the program evaluator(s) of the results while evaluation seeks to report to the superior authorities. In assessment, success means how well students progressed and failure implies how poorly the teacher performed while success in evaluation indicates how effectively the program has been managed, and failure is implicitly ascribed to the ineffectiveness of instruction. What makes the issue worth noting here is that the teacher is to blame for failure in both cases. Data in assessment are collected by concentrating on students' moment-by-moment performance in the classrooms, "emanating from alternative activities" (Genesee, 2001: 149) while evaluation involves the gathering of data by focusing on teaching performance and learning outcomes. To make decisions, assessment uses informal instruments of data collection, but in evaluation, formal instruments of data collection is employed.

- Assessment and Testing

Assessment and tests are both forms of measuring student's language learning ability but differ in many respects. *Tests* refer to specific instruments that measure the achievement and proficiency of students, and *assessment* refers to a more general concept of scrutinizing the students' learning progress. Tests are a subset of assessment. They are prepared administrative procedures that occur at identified times... (Brown, 2004). Assessment is used as a broader notion for all types of measures. In the following are further instances which account for their discrepancies.

1. Whilst tests can be used as a 'bolt-on' procedure at end-points in a learning programme, assessment is integral to the whole process of teaching and learning (Hedge, 2000).
2. Tests are product-oriented, checking how much a student has achieved or progressed, but assessment is process-oriented that focuses on how well a student has progressed.
3. Tests are carried out based on norm referencing while assessment is criterion-referenced.
4. Tests make use of quantitative criteria only, but assessment uses both quantitative and qualitative criteria.
5. Tests are administered in a formal or semiformal mode while assessment calls for an informal situation.
6. Assessment is used for low-stakes purposes but tests involve high-stakes decisions because "they are likely to have a major impact on the lives of large number of individuals, or on large programs" (Bachman & Palmer, 1996: 96).
7. It is the teacher who decides about the students' progress in assessment, but tests are scored by either the teacher or by outside authorities.
8. Scoring in assessment is known to the students in order for them to be aware of their learning process while students are shorn of knowing the outcomes of their tests.
9. Judgments in assessment are made in a dynamic manner, but fixed judgments are characteristic of different types of tests.
10. Assessment aims to enhance learning that is in opposite of tests seeking to gauge pure learning.
11. In tests, both inter-rater and intra-rater reliability indices are made use of, but in assessment, only "intra-rater reliability is of constant concern to teachers" (Brown, 2004: 32).
12. There are no Pass/Fail parameters in administration of assessment in the classroom while students are doomed to either pass or fail the tests or graded as high or low proficient learners in tests.
13. Because students are assessed in terms of what they know and what they don't know in taking the tests, they are provided with convergent types of questions. On the other hand, assessment requires the students to respond to both convergent and divergent questions mostly because the students' know-how is measured.
14. Tests involve tight constraints of time while students' performance is not assessed in terms of time limits.
15. Errors in assessment are welcomed because they are seen "as signs of learning rather than of failure" (Richards & Rodgers, 2001: 110), but errors in tests result in penalty.
16. Students are never provided with any types of feedback during the time when they are taking the tests. They, on the other hand, are supplied with feedback in assessment whenever necessary. They are keen to get feedback from the teacher and act upon what they are told (Harmer, 1998).
17. Tests see their takers as candidates and examinees who are eventually risk-takers while assessment views its assessees as input-takers and output-givers.
18. The consequential validity of tests is more demanding than that of assessment is.
19. Tests are often administered rigorously, but assessment calls for an atmosphere with plenty of flexibility.
20. Testing environment tends to change from test to test, which can affect individuals' test performance, but assessment is carried out typically in regular conditions.
21. In tests, each correct response has the same value for all test-takers while, in assessment, the degree to which a question is responded correctly varies from student to student.
22. The questions in tests cannot be repeated, but a question in assessment is capable of repetition from time to time.
23. Teachers tend to interpret what students perform in the classrooms whereas tests require deterministic scoring.
24. While questions in tests are constructed in a predetermined or preplanned manner, the questions in assessment are conducted with immediate decisions.
25. Tests measure competence, but assessment gauges performance.
26. Assessment is useful mostly for internal decisions, but tests are applied to external decision-making.
27. Tests bring about much more psychological concerns than assessment.
28. Unlike the assessees in the classroom, test-takers, especially in high-stakes proficiency tests, often do not know who the test developer is.
29. Students are provided with opportunities before and after units of instruction to assess their own performance (self-assessment) (Farhady, 2003), but tests do not offer such an opportunity for test-takers.
30. The ethical issue of fairness is controlled in less fashion in assessment than in tests.
31. Unlike assessments, tests are usually constructed and administered to make decisions on the lives of large numbers of students as well as large programs.

Pedagogical background

Historically, two approaches have been adopted as to the L2 assessment: **product-oriented** and **process-oriented**. The **product-oriented** assessment refers to the outcomes of learning process, focusing mainly on *what* the student knows or can do. In other words, it aims to assess whether and how far individuals or groups have been successful (Fulcher & Davidson, 2007). Also, it should determine what is to be taught and learned (Lambert & Lines, 2000). This type of assessment, which owes its principles to the psychometric judgment of student learning, tends to measure isolated, discrete and decontextualized skills and is most appropriate for documenting proficiency or competency in a given skill, i.e., for summative purposes. Product-oriented assessment methods are employed to provide quantitative data regarding student achievement to the people at the district and state levels. The collection of these data aims to place learners in the appropriate environments. Focusing on the product in assessment calls for employing an approach that judges the performance of an individual within a group against the whole group's performance. It is a teacher-fronted and educational system-run approach.

The **process-oriented** assessment, on the other hand, shifts away from traditional methods and gives emphasis on measuring how well students do or do not do. This qualitative assessment focuses on problem-solving, decision-making, analyzing and interpreting information. It gives learners information about how they are progressing toward goals and what the next step in the learning process might be. This type of assessment which is founded on the assumptions of the post-method era of language testing and assessment, espousing the view that "alternative ways of thinking about learning and assessing learning are needed" (Hancock, 1994), is broad-based, continuous, authentic, and grounded in knowledge of literacy, requiring collaboration on the part of both students and teacher. In other words, it is an on-going process involving the student(s) and teacher into making judgments about the student's progress in language using non-conventional strategies (Hancock, 1994). The process-

oriented assessment involves procedures and steps based on which learners perform a particular task or activity. In this approach, the idea is to explore how learners learn and what type of strategies they make use of in learning. It is a student-centered approach.

-Types of Process-oriented Assessment

Process-oriented assessment is an umbrella term to refer to three types or models of assessment as *direct* assessment, *performance (-based)* assessment and *authentic* assessment.

Direct assessment of performance is the extreme form in which the examinees are put into the actual situation and their performance is evaluated. In this case, the situation is (I think *may* not be) not manipulated to the present specific tasks (Mousavi, 1999). L2 researchers reported that its origin dates back to the worldwide adoption of CLT paradigms. Direct assessment of language performance is time-consuming and therefore expensive, particularly individualized testing (Brindley, 2001).

Stiggins (1994) defines *performance-based assessment* as the use of performance criteria to determine the degree to which a student has met an achievement target. Hamp-Lyons (1997), quoting from Meisels, Dorfman, and Steele (1995), argued that performance assessment is longitudinal, has contextual objectives, monitors progress, promotes learner learning, enhances learner motivation, is instructionally relevant, informs instructional decisions, and contributes to classroom change. Performance (-based) assessment is an approach that seeks to measure pupil learning on basis of how well the learner can perform on a practical real-life task such as the ability to write a composition or carry out a small talk. In this type of assessment, students make a constructed response and engage in higher-order thinking; tasks are open-ended, meaningful, engaging, and authentic while they require an integration of language skills, and both process and product are emphasized with a focus on depth of student's language mastery rather than on breadth (Gipps, 1994). Brown and Hudson (1998) wrote:

The principal advantage of performance assessments is that they can come close to eliciting authentic communication (at least insofar as authentic communication can be elicited in any testing situation). Advocates of performance assessments maintain that performance assessments provide more valid (a) measures of students' abilities to respond to real-life language tasks, (b) estimates of students' true language abilities than traditional standardized multiple-choice assessments, and (c) predictions of students' future performances in real-life language situations.(p. 662).

Authentic assessment can be described as procedures for evaluating learning performance using multiple forms of assessment that reflect student learning, achievement, motivation, and attitudes on instructionally-relevant classroom activities (O'Malley & Chamot, 1999). It is a term used largely in the USA where the intention is to design assessment which moves away from the standardized, multiple-choice type test towards approaches where the assessment task closely matches the desired performance and takes place in an authentic, or classroom, context (Gipps, 1994: 11). Thus, assessment is authentic when it corresponds to and mirrors good classroom instruction and when the results can be used to improve instruction based on accurate knowledge of student progress (O'Malley & Valdez Pierce, 1996). It implies that tasks used in assessment are valued in the real world by students. Alternative methods of assessment such as self-assessment, teacher assessment, collaborative assessment, and portfolio assessment are examples of authentic assessment (Richards & Renandya, 2002).

-Which is the current focus and why?

Worthen (1993) claimed that although direct assessment, authentic assessment, and performance (-based) assessment have subtle variations in emphasis, the terms are often used interchangeably as all of them have common characteristics. The reason is that all the three types focus on students' processes, products, or performances. But there seems to be a third argument. Borrowing from Widdowson's (1979: 165) view conceiving of authenticity as an attribute 'bestowed' on texts by a given audience, authenticity in testing terms is not an assessment type but an attribute to or a property of all types of alternative assessment or which is disclosed by the assessors in assessment types. This is the salient trait of authentic assessment. However, authentic assessment represents the real-life picture of learning. In this case, students are asked to demonstrate their knowledge, skills, or competencies in whatever way they find appropriate (Guba & Lincoln, 1989).

It was not until the late 1970s that 'authenticity' gained its greatest popularity under the penetration of process-oriented theories and principles into communicative language teaching and testing, threatening the kingdom of the decontextualized, one-shot methods of measuring students' learning. The notion has been grounded on the premise that teachers can assess students' real performances in relation to real contexts. The very rationale has brought about many forms of assessment, as well. Authentic assessment forms, therefore, serve two functions: they "provide teachers with useful information that can form the basis for improving their instructional plans and practices" in one hand, and they "provide students with a tool to be more involved in their learning, and give them a better sense of control for their own learning", on the other (Richards & Renandya, 2002: 335). According to O'Malley and Valdez Pierce (1996: x), assessment is authentic when it corresponds to and mirrors good classroom instruction and when the results can be used to improve instruction based on accurate knowledge of student progress. These authors used the term 'authentic' to describe the multiple forms of assessment that are consistent with classroom goals, curricula and instruction (1996: 2). Therefore, such an assessment places greater demands on teachers than on the application of one-shot tests. Wiggins (1989) also offered two criteria of assessment authenticity: one that reflects the challenges, work, and standards engaging practicing professionals; and the other which involves a student with opportunities for dialogues, explanations, and inquiry.

The growing interest in adopting authentic measures of assessment in second or foreign language classes indicates that they are of salient characteristics. Here are a number of features that have been proposed by the researchers and scholars in testing and assessment.

“ They encourage students to use declarative knowledge while studying (O'Malley & Chamot, 1999).

“ They support higher-order of thinking skills (Shohamy, 1994).

- “ They focus on using creative approaches to discover what students know and can do.
- “ They enable L2 teachers to learn more about their students' strengths and weaknesses in using the four language skills of reading, writing, speaking, and listening (Shohamy, 1994).
- “ They describe meaningful tasks that require students to perform and produce knowledge rather than simply reproduce information others have discovered (Wiggins, 1993).
- “ They present the student with the full array of tasks that mirror the priorities and challenges found in the best instructional activities (Wiggins, 1993).
- “ They allow for a thorough analysis of a student's performance (Shohamy, 1994; Valencia, Hiebert & Affleberg, 1994; Valette, 1994).
- “ They provide parents and community members with directly observable products and understandable evidence concerning the students' performance (Bryant & Timmins, 2002).
- “ They present students with real-world challenges that require them to apply their relevant skills and knowledge (Guba & Lincoln, 1989).
- “ They achieve validity and reliability by emphasizing and standardizing the appropriate criteria for scoring students' products (Wiggins, 1993).
- “ They call upon teachers to perform new instructional and assessment roles (Brown, 2004).
- “ They maximize learners' self-confidence and minimize their psychological concerns.
- “ They require students to perform and produce knowledge rather than simply reproduce information others have discovered (Wiggins, 1993).

Authentic assessment focuses more on measuring learners' ability to use language holistically in real-life situations and is typically carried out continuously over a period of time (Richards & Renandya, 2002). Therefore, two points of note should be taken into account in this regard: how to plan and manage the assessment process in the course of time, and how to judge student leaning and performances so as to mirror his/her actual progress. Although authentic assessment is costly and time-consuming (Wiggins, 1989, 1993; O'Malley & Chamot, 1999; Worthen, 1993), it will take effect if it is planned carefully. Shohamy (1994) introduced several types of authentic assessment which include observations, questionnaires, interviews, homework, and documents from test-taker files, letters, learner diaries, portfolios, and self-assessment, but contemporary research studies seem to have identified many other forms. Such alternatives as project work, journals, peer assessments, student-teacher conferences, homework, cooperative test constructions, checklists, peer-editing, exhibitions, records of classroom participations, progress grids, learning logs, note-taking, anecdotal records, audiotapes of discussions, proof-reading, drawings, videos of role-plays, work samples, protocol analysis are all considered by most educators as forms of authentic assessment.

Issues of Assessment Quality

Literature on assessment has revealed that in order for it to be successfully implemented in L2 language classrooms, four quality-oriented criteria should be regarded as vital. Therefore, a classroom assessment should at least contain *reliability* and *validity*, which, as Genesee & Upshur, (1996: 56) claimed, are both critical for judging the quality of qualitative information, and *fairness* that should also be highly appreciated in design and use of assessment. Washback is the fourth issue in guaranteeing the quality of assessment in L2 classrooms.

- Validity Issue

Validity, especially in alternative assessments, is the extent to which an assessment measures what it purports to measure (Gipps, 1994). In other words, it is the extent to which the information you collect actually reflects the characteristic or attribute you want to know about (Genesee & Upshur, 1996: 62). Moreover, once proposed by Messick (1989), it is now recognized as being a *unitary* or *unified* concept (Lambert & Lines, 2000; Kunnan, 2005) focusing on a backward as well as forward use. In classroom assessment, at least three main types of validity should be considered. The first is the *content validity*, which is the correspondence between curriculum objectives and the objectives being assessed. To put it another way, it is an aspect of construct validity that emphasizes evidence bearing on the appropriateness of the knowledge, skills, and abilities measured by an assessment (ETS, 2002: 68). Since assessment in the ESL classrooms based on the contemporary paradigms are supposed to benefit from a large quantity of alternative forms, research to date has shown that they have more or less a high degree of validity in content for two major reasons: one is that they to a considerable extent adhere to course objectives, and the second is their capability of integrating instruction and assessment. Therefore, the issue of content validity decreases the after-effect impact in teaching and learning if the domain is properly appreciated. *Consequential validity* is the second type of validity, referring to the way in which the assessment is used to benefit teaching and learning processes and to benefit students (O'Malley & Valdez Pierce, 1996). Assessment forms have *consequential* validity to the extent that they lead teachers to focusing on classroom activities which support students learning and are responsive to individual needs (1996: 26). Messick (1989), McNamara (2000) and some other scholars claim that the consequences of an assessment are potentially important, because the focus is on the impact it can have on variables at the state and district levels. It also includes test washback on teaching and learning (Fulcher & Davidson, 2007). The third type of validity is the *Ipsative validity* which is used when teachers take into account their students' performance that is formatively assessed during the classes, not make increasing use of their past papers or performance as a valid criterion to judge their learning abilities. Put simply, the pupil evaluates his/her performance against his/her previous performance (Gipps, 1994). Lambert and Lines (2000) regard this type of validity as important because it places the student at the center of the assessment activity and provides diagnostic information on the progress (or otherwise) of the individual. It is also called ' *pupil-referenced*' validity.

- Reliability Issue

Genesee and Upshur (1996) define 'reliability' as the freedom of nonsystematic fluctuation, or the degree to which an assessment yields consistent results. An assessment is reliable when there is little difference in learners' scores or in judges' ratings across different occasions or different judges (Brindley, 2003). Therefore, reliability needs to be based on performance instead of idiosyncratic scores that have no preset criteria (O'Malley & Valdez Pierce 1996: 20). William (1992) calls it *dependability* (cited in Lambert & Lines, 2000: p. 11). By this, he argues that classroom assessment is dependable when *disclosure* (i.e., the extent to which a student gets a question right or wrong depending on the nature of the question itself), and *fidelity* (i.e., the way the evidence is, or is not recorded) are present (Lambert & Lines, 2000: p. 11). Three general sources of problems have been found with the reliability of a classroom-based assessment: fluctuations in scoring people gathering the information; the instrumentation of data collection; and fluctuations in the learner. Genesee and Upshur (1996: 59) propose some very practical ways of improving the reliability. One is that the raters or assessors should be trained or experienced so that they know exactly how to get the desired information. These authors suggest using more than a single observer, interviewer, or composition reader; certain scholars argue against the idea and say, "Since classroom tests and assessment rarely involve two scorers, inter-rater reliability is seldom an issue. Instead, intra-rater reliability is of constant concern to teachers" (Brown, 2004: 32). The second solution is to employ students' performance on different occasions for making decisions about them although subjectivity in scoring *may* produce unreliable measurement (Henning, 1987) (emphasis added); and to avoid this threat, the raters should use a variety of methods of information collection in order to remove the bias or inaccuracy of using one method.

- Fairness Issue

Kunnan (2005) emphasizes that the most important challenge in large-scale assessment is the issue of fairness. Fairness can be defined as treating all individuals equally and giving all individuals an equal opportunity to contribute to the research process or, in the case of assessment research, to demonstrate their ability (Lynch, 2001). Gipps (1991) calls it 'equity' in the design and use of assessment (Cameron, 2001). Messick (1994) argues that issues of fairness are at the heart of performance assessment validity. Therefore, all students taking performance assessments should have reasonable opportunities to manifest their know-how without difficulty. Fairness will also need to address the consequences of assessment; that is, we need to examine the uses to which our assessment procedures are being put and the intended as well as unintended effects on the individuals being assessed (Lynch, 2001: 232). Emphasizing the impartiality of assessments (or tests), Brown (2005) felt that teachers would generally like to ensure that their personal feelings do not interfere with fair assessment of the students or bias the assignment of scores. Fairness, or equity as Gipps suggested, principles require that students be given plenty of chances to show what they do, and that their language learning be assessed through multiple methods (Cameron, 2001). Cameron further stated that fairness is important in planning and designing assessments that the content is scrutinized to make sure that culturally unfamiliar pictures or concepts do not reduce children's chances to demonstrate their language learning (2001: 226). Fairness in assessment starts with fairness in the learning process (Kimball & Hanley, 1998). To them, students are given an opportunity to examine program standards and outcomes and assessment criteria at the beginning of their program of study, and a midyear review is conducted to evaluate student standing and level of performance against the specified standards. Thus, students have a chance to learn and practice their skills during internship. From the fairness perspectives, Lambert and Lines (2000) indicate, norm referencing intuitively suggests unfairness; being judged against a group of other people depends on the make-up and qualities of that group, and criterion referencing, on the other hand, suggests equality of opportunity. Fairness is believed to have a relationship with 'power'. In language assessment, *power* means "who decides what will be done to whom" (Herron, 1988; see in Lynch, 2001: 232). It includes forms and relations. Foucault (1982) categorized the basic forms of 'power' as 'domination', 'exploitation', and 'subjection' (see also in Lynch, 2001: 232). Accordingly, what matters here is to claim that whenever the power comes in, the fairness goes away. Fairness is not without shortcomings in the authentic assessments. O'Malley and Valdez Pierce (1996: 28) identified four problems with fairness. One problem is that the performance called for in authentic assessment forms is often highly language-dependent, either oral or written. Second, the responses called for in performance assessments involve complex thinking skills. The third problem with fairness is that authentic assessments are often used to measure student knowledge in depth in a particular area. And finally, the use of authentic assessments might exacerbate the problems with culturally unfamiliar content. If the content related to the single theme is unfamiliar, the student may be unable to respond to any of the questions contained in the assessment.

-Washback issue

Washback is an aspect of impact, or a facet of consequential validity, which has become a major area of study within the applied linguistics, especially language testing and assessment. It is sometimes referred to as *backwash* in general education (Hughes, 1989, 2003; Heyneman, 1987; Fullilove, 1992; Spolsky, 1994, 1995; Biggs, 1995, 1996), *measurement-driven instruction* (Popham, et al. 1985; Popham, 1987), *Curriculum alignment* (Shepard, 1991b, 1993), *systemic validity* (Frederiksen & Collins, 1989; Pierce, 1992; Berry, 1994; Cohen, 1994) or *test impact* (Bachman & Palmer, 1996). Washback and impact were once used synonymously and interchangeably. Now, they are, though related, definitely different in scope. "More often impact is taken to be the superordinate while washback refers to the narrower situation of the language classroom." (Davies, 2003: 361). In other words, impact covers a broader area of education, which is "the wider effect of tests on the community as a whole" (McNamara, 2000: 74). "Washback occurs when it is the testing instrument rather than the statement of desired learner outcomes that determines the nature of the curriculum and the course of instruction" (Valette, 1994: 10). Achievement Tests, national entrance examinations, TOEFL tests, IELTS tests, and the Oral Proficiency Interview all exert washback effects.

Washback has come to be known as one of the ethical issues of high-stakes assessments since 1990s following similar concerns generated by Messick's (1989) expanded framework of validity (Bachman, 2000). Since then, scores of studies, theoretically or empirically, have been done on the issue. Different views and arguments have also been raised as to the concept of washback. But, more or less, it has been generally and widely accepted that high-stakes tests and examinations result in a change in teaching and learning behavior. What makes the authors differ is a number of hypothetical questions they have raised about the *nature* of the test effect (Shohamy, 1993a; Alderson & Wall, 1993), the *way* a test affects (Cohen, 1994), the *areas* of test effect (Wall & Alderson, 1993), the *degree* of test effect (Shohamy, Donitsa-Schmidt, & Ferman, 1997), the *essence* of test effect (Cheng & Curtis, 2004; Alderson, 2004), the *existence* of test effect (Alderson & Wall, 1993; Shohamy, 1993a), the *importance* of test effect (Alderson & Wall, 1993), the *scope* of test effect (Wall, 1997), the *person* a test affects (Alderson, 1992; Watanabe, 2004; Cheng & Curtis, 2004), (the *exact nature* or *mechanism* of test effect (Bailey, 1996), the *intensity* of test effect (Shohamy, et al. 1997; Cheng, 1998a), the *duration* of test effect (Watanabe, 2004), the *cause* of test effect (Alderson, 2004), the *value* of test effect (Wall & Alderson, 1993; Messick, 1996), the *intentionality* of test effect (Messick, 1989; McNamara, 1996), and the *form* of test effect (Alderson, 2004).

Although every concept related to washback as mentioned in the above has been investigated, or will be explored in the future, much of the research to date on the influence of high-stakes testing have addressed the 'value' of the washback phenomenon — whether there is a positive or a negative washback, or in Hughes's (1989, 2003) and Buck's (1988) terminology, a *beneficial* or a *harmful* effect, on teaching and learning (emphasis added). Wall and Alderson (1993), as pioneers of washback studies, for example, argued that "...tests can be powerful determiners, both positively and negatively, of what happens in classrooms" (p. 41). It is worth noting that before

this time, the washback was largely viewed as negative (Fulcher & Davidson, 2007). The assumption underlying the research into the presently-understood washback is that if the teaching is to be to the test in which learners are provided with test-taking skills and strategies or test-driven activities or with a need for just passive acquisition, it is claimed to have negative or harmful impact on students' learning and teaching. On the other hand, when the teachers are in pursuit of assigning to students the activities which develop knowledge-driven, cognitive and metacognitive skills and strategies, the washback will place a positive effect on learning as well as teaching.

Most of the findings from the empirical studies of washback in different countries have confirmed both beneficial and harmful effects although there were some taking no sides with any, or else. Smith (1991b), for example, found that teaching to the test narrows curricular offerings and modes of instruction, reduces time of instruction, decreases teachers' capabilities to teach content and to use materials incompatible with high-stakes testing. Anderson et al. (1990) also reported that test-driven instruction limits the topics to be assessed, minimizes students' critical thinking, and encourages students to adopt the memorization approach. Alderson and Hamp-Lyons (1996) argued that teaching to the TOEFL test encourages teachers to talk more, gives rise to the use of meta-language, and reduces the opportunities for the mutual interaction, emphasizing that the effect varies in degree and type among teachers. From the positive side of washback, researchers found that tests will offer more beneficial effect if the accountability in testing increases. 'Creative and innovative testing....can, quite successfully, attract to itself a syllabus change or a new syllabus...' (Davies, 1985; cited in Cheng & Curtis, 2004: 10-11). "Positive washback benefits teachers, students, and administrators because it assumes that testing and curriculum design are both based on clear course outcomes that are known to both students and teachers/testers" (Coombe, Folse & Hubley, 2007: xxv). "To achieve positive washback effect, Hasselgren (2000) maintained, the tasks should promote good learning as well as assessment activities [in which] both pupils and teachers should develop their ability to assess, based on explicit criteria" (p. 262). Arguing that well-designed performance assessments can actually provide strong positive washback effects, especially if they are directly linked to a particular curriculum, Brown and Hudson (1998) wrote:

A positive washback effect occurs when the assessment procedures correspond to the course goals and objectives. For instance, if a program sets a series of communicative performance objectives and tests the students using performance assessments (e.g., role plays, interviews) and personal-response assessments (e.g., self-assessments, conferences), a powerful and positive washback effect can be created in favor of the communicative performance objectives. Positive washback occurs when the tests measure the same types of materials and skills that are described in the objectives and taught in the courses (p. 668).

But there is now a widespread claim that whether tests promote or inhibit teaching and learning depends on a number of such contextual factors as the teachers' educational background, the inadequate communication of information by educational authorities, the lack of professional development support, the immediate motivation of learners, the frequently unpredictable ways of classroom interaction development, the teachers' different beliefs about effective teaching, the timing of observations relative to the test date, local conditions in classrooms, the low levels of teacher awareness, the established traditions of teaching, research methodology, etc (Alderson & Wall, 1993: 116; also see in McNamara, 2000: 74). This is true because of a generally accepted assumption that washback is a complex phenomenon in the current educational system (Bachman & Palmer, 1996; Alderson & Wall, 1993; Watanabe, 1992, 1996a, 1996b, 2004; Alderson & Banerjee, 2001; Alderson & Hamp-Lyons, 1996; Shohamy, et al. 1997; Wall, 1996, 2000; Bachman, 2000; Davies, 2003; Alderson, 2004; Cheng, 2004).

Unfortunately, there is not much evidenced and empirical information about how and to what degree the washback has impact on L2 learning and teaching in low-stakes situations. Most of the research that has been conducted concerned the kind and degree of washback and identifying its factors affecting learning and teaching in high-stakes testing. This is due to the complexities in the educational systems at both micro- and macro-levels, the complex relationships between contextual factors and instructional methods as well as learning, and the complex characteristics of the performance or alternative assessment forms. Also, the fundamental difference between language testing practices and classroom assessment practices actually resides in the variety of assessment types that, at least, teachers use (Brown & Hudson, 1998). This can be seen as a good indication of the laborious undertaking, or limitations, of doing washback research on the classroom performance-based or authentic assessment forms in order to find out how and to what degree these assessment practices affect the instruction and learning, or other stakeholders.

Apart from the negative and positive impact of washback in large-scale assessments as discussed in the above, Brown (2004) suggests a form of washback that occurs more in classroom assessments. He conceives of it as the information that "washes back" to students in the form of useful diagnoses of strengths and weaknesses (2004: 29). In this case, teachers can employ a variety of alternatives to assessment in order to entail high-order thinking or cognitive processing to develop the awareness, perception, integration and application of the context knowledge for the purpose of engaging the students into an interactive atmosphere of accomplishing the given problem-solving activities. This is a guaranteed means of achieving a positive washback because teachers are underscoring the idea of what pupils usually *can* do, instead of documenting what pupils *should* be able to do (Hasselgren, 2000; Wiggins, 1989a, 1993). McNamara (2000), describing performance assessments as having presumed positive or better washback (assumed to be more progressive) than individual item formats, preparation for such assessments will presumably encourage teachers and students to spend time engaged in performance of realistic tasks as part of the teaching (p. 74). This will not take effect unless a reform in curriculum design as well as a change in the educational policy is made because assessment reform unequivocally calls for curriculum reform or innovation (Andrews, 2004).

Confirming that washback effect is also commonly found at the classroom level Valette (1994: 11) suggests that the creative classroom teachers can use the washback effect to great advantage through:

Ø Establishing a grading system that fosters the type of language learning which will promote the desired learner outcomes.

Ø Developing tests in a format that encourages learners to develop desired study habits.

Ø Watching for other learner outcomes, using formal grading criteria to promote positive washback, that are not formally measured but nonetheless observable and thereby conducting an on-going informal assessment of how well students are progressing.

Although achievement of positive washback in the reform paradigm calls for the use of authentic texts and task or, strictly speaking, the use of alternative assessment

(Bailey, 1996), Hamp-Lyons (1997) reported a need for further studies of the impact of alternative assessments on the same basis that we apply to traditional forms of assessment. He added, "We cannot assume that because alternative assessments start from humanistic concerns, they produce outcomes that do only good and no harm (1997; p. 297)."

Purposes of Assessment

A review of the literature in the contemporary approaches to L2 assessment shows that assessment, no matter incidentally or intentionally, serves certain purposes. It aims to find answers to three questions of how assessment is carried out, what information is collected, and how the collected information is interpreted in a pedagogical setting. Therefore, teachers make plans for them from the very beginning of the study term. Therefore, I have classified the purposes of assessment as **global** and **local** objectives for better understanding. Global objectives address the decisions made for the wider community within an educational system. They occur outside the classroom context. They also target the decisions *about* students' learning or competence. On the other hand, the assessment is local when it is carried out in the classroom context in which the teacher is in pursuit of doing the tasks or activities to foster learning. All the objectives at this level focus on the decisions made *for* students' learning or performance.

Assessment at the **global** level is used:

1. to place learners in the programs (i.e., placement), or to determine whether learners have adequate language proficiency (i.e., selection) before the course begins. Bachman (1990) refers to the former as 'identifying the appropriate instructional levels', and the latter as 'an admission decision'.
2. to identify learners' strengths and weaknesses with the aim of either making revisions of the given curricula/materials, or introducing remedial activities (i.e., diagnosis). In other words, it should elicit information on what students need to work on in the future (Brown, 2004: 47).
3. to measure the effectiveness or success of instructions to modify or redesign them to meet the needs of students, or that of programs at the school, district, state, levels (i.e., evaluation). The outcome of assessment can inform the evaluation and improvement of courses and programs (Cameron, 2001: 220), but Genesee and Upshur (1996: 14) argue that L2 evaluation in the classroom is concerned primarily with improving instruction that student learning is enhanced.
4. to report students' progress and achievement to supervisors, administrators, stakeholders, macro-/micro-policy-makers, funding authorities through gathering and documenting systematic information (i.e., accountability). An aspect of accountability is the question of determining norms of language behavior which will act as a reference point in the assessment (McNamara, 2000: 72).
5. to help teachers make academic decisions about individual students' achievement regarding course goals and mastery of course content (i.e., academic proficiency).
6. to provide the means for selecting by qualification (i.e., certification) (Lambert & Lines, 2000: 4). For a certification to take place, summative assessment provides a baseline for those who wish to proceed to the next stage in their education or towards a career (2000: 90).

Assessment at the **local** level is used:

1. to develop higher-order thinking skills
2. to foster intrinsic motivation in students
3. to treat students with respect and fairness
4. to promote students' autonomy
5. to have students employ more learning and use strategies
6. to encourage students to construct their own classroom tasks
7. to elicit students' low-stakes performance
8. to have students check their own errors themselves
9. to require student-teacher and student-student collaboration
10. to encourage self-regulated learning
11. to increase students' responsibility for learning
12. to trim down the negative washback effect
13. to enhance students' self-esteem
14. to lead students to the state of being strategic learners
15. to involve students into self-assessment
16. to reduce the effect of methodological constraints
17. to minimize learning anxiety

18. to reflect growth over time
19. to welcome errors to diagnose learning difficulties
20. to foster the acquisitional learning
21. to improve students' know-how
22. to develop learning literacy
23. to find out who the students really are
24. to have teachers teach students and not the subject matter
25. to engage students into doing creative activities
26. to acquaint students with how to take risks in learning
27. to move teachers toward quality
28. to reduce teacher's role to a facilitator and coordinator
29. to lead students to having a sense of pleasure
30. to induce a collaborative setting of leaning goals
31. to get teachers to be responsive to students' developmental needs

Classroom Assessment Design

Another issue of the L2 classroom assessment is how "to design accountability assessment which will provide good quality information about students' actual performances without distorting good teaching (and therefore learning) practice" (Gipps, 1994; 175). In fact, the design of an effective assessment paves the way for beneficial washback. Assessment design requires teachers to have determined the appropriate objectives as in the above, and they should also take great care that the design does not affect or distort those objectives. Therefore, assessment in the classroom should be carried out through six specific steps followed in order of function. These steps include *planning*, *data collection*, *data organization*, *data evaluation* and *final reporting*.

1. Planning

Planning is the first and fundamental step in L2 classroom assessment . It determines *why* a teacher should ever assess, *what* he/she should measure, and *who* he/she should evaluate. It also specifies *how* the assessment process should be carried out. Moreover, there should be a focus on the *time* of assessment administration and the *degree* to which a teacher can assess the students.

Ü Why to assess

Teachers have meaningful goals for instruction and clear purposes for assessment (Wolf, 1993). They assess to meet the course goal, to check the effectiveness of on-going instruction, and to improve the learning process. One of the great purposes of assessment is to gather information in order to modify and adapt the materials and curricula. They make decisions about students' current learning needs (Genesee & Upshur, 1996). In short, teachers assess their students to accomplish the objectives discussed mostly at the *local* level in the above.

Ü What to assess

Teachers cannot assess whatever they themselves like. In classroom assessment, teachers are supposed to assess students' current abilities in a given skill or task. The teacher can assess students' oral communication ability if they need to communicate orally. If they are going to develop their academic skills, the teacher can assess how and how far they have mastered them. Teachers may also assess two or multiple abilities of their students at the same time. They may evaluate their pupils' vocabulary build-up as well as how well they have progressed in reading comprehension, and also how far they have observed the suprasegmental rules in reading passages. Thus, it is urgently required that teachers should have adequate perception of what they assess in the classrooms.

Ü Who to assess

It may seem strange to ask whom a teacher should assess in the classroom, but the issue is of great concern. Teachers should treat students as 'real learners', not as course or unit coverers. Teachers should be aware that students enjoy competence and proficiency of different levels and types or different schematic and systemic knowledge. They should also predict that some students are more active and some are less active; some are quick at learning and some are slow at it. In addition, they should expect that "students vary in how they learn by establishing different goals for learning" (O'Malley & Valdez Pierce, 1996: 10). Therefore, classroom assessment

calls for a prior realistic appraisal of the individuals teachers are going to assess.

Ü How to assess

Teachers employ different instruments, formal or informal, to assess their students. Brown and Hudson (1998) reported that teachers use three sorts of assessment methods – selected-response assessments, constructed-response assessments, and personal-response assessments – to evaluate students' ability or progress in L2 learning. They can adjust the assessment types to what they are going to assess. Because of their difficulty in construction and failure of eliciting any productive language as well as their high-stakes impact, the selected-response assessments at least are not much favored in the formative classrooms. Therefore, students are required to produce language through composing written and oral discourse and create and/or enhance the real-world communication. This is not true that the other two methods are free from shortcomings, but they are more in line with the formative assessment paradigms. Here, I suggest using the convergent and divergent approaches to L2 assessment in order not to commit misconceptualization. As the terms imply, students are engaged into doing the higher-order thinking skills. This will cause learners to adopt the self-assessment and self-correction practices in order to achieve their autonomy. What matters here is that the assessment should be carried out *with* learners, not to them because, as Harmer (2001) states, students need to know *how* and for what they are being assessed.

Ü When to assess

Contemporary research into L2 assessment has indicated that assessment is interwoven into instruction. Whatever teachers assess, they are actually in the process of teaching. They assess when they are going to make instructional decisions at the formative and summative levels, even if those decisions are small. For example, they assess when there is a shift in instructional units; when there is a check and recheck of the students' ability in perception and production of learning items at different intervals; when the effect of the given materials or curriculum on the improvement of learning process is examined; and when they intend to fine-tune their teaching process to the learning objectives.

Ü How much to assess

There is no touchstone to weigh the degree to which a teacher should assess students while the classroom instruction is believed to be interlocked with the formative assessment types. But it doesn't mean that teachers can evaluate their students to the extent that they prefer. It should be predicted that students differ in styles, types of competence, ability to strategy learning and use, views toward learning, background knowledge, psychological and social as well as personality factors, etc, so that assessment should be limited to every individual's needs, ability and knowledge. Although such factors are all hypothetical and to some extent economically incapable of being evaluated before or at the beginning of the course, prediction as part of planning can at least prevent teachers from overassessment or underassessment.

2. Collection

In order to make decisions about the effectiveness of assessment in different aspects of pedagogy, we need to gather adequate and appropriate information. In the past, final tests were the only devices used as the end-of-the-term instruments to collect data about the students' learning achievement and progress. They are still of great interest to many language institutes, teachers, administrators as well as micro- and macro-policymakers being aware of their potential drawbacks in authentically assessing students' learning. The reform in evaluation based on the process-oriented principles involved adopting diverse alternatives to L2 classroom assessment. Genesee and Hamayan (1991) introduced three ways which teachers use to keep assessment records in ESL classrooms. They are student portfolios, narrative records, and checklists. But the more recent studies suggest further cases in this regard. From among the miscellaneous alternatives, teachers are supposed to gather information through such most common types as student portfolios, journals, narrative records, quizzes, checklists, drawings, student-teacher conferences, peer-editings, self-assessments, peer assessments. They employ them because these alternatives are believed to reflect the real-world and authentic picture of the learning progress. Record keeping and collections of work samples by both teachers and students provide systematic information that facilitates communication (Tierney, Carter & Desai, 1991).

3. Organization

The collected information cannot by itself suffice to be evidences based on which decisions are made. They should be systematic and well-structured. Systematic collections need to be carefully planned, just like instruction (O'Malley & Valdez Pierce, 1996). Teachers should, therefore, organize the information they garner through various assessment types in order to make possible the evaluation and reporting processes and help the authorities to make easier decisions.

4. Selection

Although structuring, organizing and sequencing information are at the service of facilitating optimal processing, not all the collected and organized information is needed for the immediate reporting. Teachers should select those pieces of assessment data fitting the teachers' and students' immediate needs and being of use for the next classroom activities and materials designing.

5. Evaluation

Before reporting to the people in different levels, judgments should be made regarding the effectiveness of instruction and curriculum and how much the students have been able to progress toward the goals they have set. This evaluation process should preferably occur after each lesson, each instructional session, or each chapter until the end of the term. Examining student progress at the end of major units can help one decide whether the students are ready to proceed to the next unit and for planning for next unit (Genesee & Upshur, 1996). Assessment at the end of each unit of instruction can also provide information about how effective the unit was (1996: 49).

6. Reporting

The final step of designing classroom assessment is to report the assessment information to certain people. Therefore, there are five audiences in this regard, each of

whom need their own share of them. Teachers, students, administrators, parents, and stakeholders/educational policy-makers are the people who receive the assessment information or results.

à **Teachers**

The main users of the information are certainly the teachers themselves. They use them to check the effectiveness of instruction and course materials. They also make decisions about students' needs for the upcoming term. What is of great note to teachers is to know how well their students could reach their stated goals. They, therefore, evaluate student progress or achievement. Teachers have a greater objective; that is, they use the information to do more careful planning for the next instructions

à **Students**

Because the assessment is said to be conducted *with* the students, they can make full use of the assessment data or results. As stated in the above, involving students into assessment can give them clear feedback as to their progress in language proficiency as well as make them more accountable for their learning. Students (can) reflect on what they have learned. They can take more active roles in making decisions about what they need for the next classroom learning. Reporting to students can also facilitate the next instruction process because they will realize what the instructional objectives are, while these objectives are set *with* them. In this case, students' learning awareness will increase considerably thanks to their growing assessment awareness.

à **Parents**

The process-oriented assessment will not put the parents out of the educational program, though they are the out-of-class audiences. Valdez Pierce and Gottlieb (1994) regard the parents as at least home collaborators and contributors in the portfolio process. Information reported to them can provide the parents with clear feedback and concrete evidence of their children progress in language proficiency. Parents can use the information to monitor and supervise their children's work and assignments at home based on the suggestive directions given by the teacher. They can also supply the teacher with information they collect about their children's learning appropriate for internal decision-making. Of course, the parents' knowledge plays a crucial role in this respect. Therefore, reporting the assessment data to parents can create a communication line between them and teachers so that they can both monitor the student learning more effectively through exchanging views. This exchange may be made in letters, phone calls, or through electronic devices. This happens when the students are young, no matter what level of proficiency they are at, because adult students, with high or low proficiency, or sometimes even young learners with high proficiency, welcome less or no support from their parents.

à **Administrative authorities**

Because teaching and assessment occur within the framework of educational systems, the information obtained should be reported to different administrative authorities for making their own decisions. Therefore, teachers are accountable to the educational administrators as well, although educational authorities may collect the accountability information in some cases (Genesee & Upshur, 1996). Typically, Stern (1983) classified the educational systems into two levels: local (i.e., school administration) and regional (i.e., Higher Education). The tasks of administration are varied (1983: 427).

▷ **School administrators**

School administrators are also the outside-classroom audiences who need reports to make a variety of decisions thereupon about the related issues at more global levels. Using them, they deal with more convenient and careful scheduling and curriculum planning. They focus on the curriculum which is to be articulated both horizontally and vertically, that is, setting and attaining goals at global and specific levels, respectively. They also wish to make sound decisions about different needs for different proficiency levels as well as the issue of inclusion, exclusion, or modification/revision of the given materials in line with the program policy. Placement the students into the correct levels/classes is another decision that such administrators make accordingly. Their major concern would be the administrative accountability; i.e., maintaining and enhancing the quality of institutional programs.

▷ **Higher Educational administrators**

Part of the assessment information should finally go to the box of the people who "ensure the proper functioning of the entire system and of each institution" (Stern, 1983: 428). In other words, they make decisions about whether the whole educational program could meet the educational policies at the global level. They also use the assessment data together with their own evaluative evidences to demonstrate to what extent they should provide the wished-for facilities as well as financial, political and managerial supports to the language schools, teachers, teacher developers, etc.

Classroom Assessment Design

image2

It is worth noting that the assessment information the audiences receive differs in amount and type. For example, teachers need the data which are instruction-specific and detailed; that is, they use part of the assessment results that determine how to develop the teaching capabilities for the sake of learning enhancement. Students, on the other hand, receive the learning-specific information, but it is not the same amount as the teachers' or the others'. The data needed here should influence the learning accountability. Parents, school and regional administrators in the field of education reflect on the broader aspects of the assessment results; that is, the public or external accountability is taken into account. What makes the issue more noteworthy is that it is the teachers who take the furthest amount of the assessment data, and the administrative sections receive the least because they themselves are evaluators and rely on their own evaluative judgments as well.

Principles of L2 Classroom Assessment

Principles are a set of beliefs and theories or conceptual frameworks developed by applied linguists for practical understandings. Teachers employ them in effective approaches to L2/FL teaching and learning. From the testing and assessment perspectives, teachers need certain assumptions relevant to the assessment issues in order to make more effective decisions about their students' actual learning abilities. As a matter of fact, these principles should be addressed in pre-service and in-service teacher education courses which can definitely raise the learner teachers' awareness of the assessment knowledge in their future actual practice. In this paper, teachers are provided with fourteen principles of L2 classroom assessment as follows.

1. Assessment should be formative.

Classroom assessment should be carried out formatively in order to inform on-going teaching and learning. It should be formative because it refers to the formation of a concept or process. This can be done through providing learners with timely and immediate feedback, the feedback having the feature of processibility, resulting in enhancing the learning potential of the pupils (Lambert & Lines, 2000). To be formative, assessment is concerned with the way the student develops, or forms (2000: 14). So it should be *for* learning. It is also designed to stimulate growth, change and improvement in teaching through reflective practice. In other words, it has a crucial role in "informing the teacher about how much the learners as a group, and how much individuals within that group, have understood about what has been learned or still needs learning as well as the suitability of their classroom activities, thus providing feedback on their teaching and informing planning " (Rea-Dickins, 2001: 457). Teachers use it to see how far learners have mastered what they should have learned. So classroom assessment needs fully to reach its formative potential if a teacher is to be truly effective in teaching.

2. Assessment should determine planning.

Classroom assessment should help teachers plan for future work. First, teachers should identify the purposes for assessment – that is, specify the kinds of decisions teachers want to make as a result of assessment. Second, they should gather information related to the decisions they have made. Next, they interpret the collected information—that is, it must be contextualized before it is meaningful. Finally, they should make the final, or the professional, decisions. The plans present a means for realizing instructional objectives which are put into practice as classroom assessment to achieve the actual outcomes (Genesee & Upshur, 1996).

3. Assessment should serve teaching.

Classroom assessment serves teaching through providing feedback on pupils' learning that would make the next teaching event more effective, in a positive, upwards direct. Teaching and learning dictate the form and timing of assessment. It is a fact of life for teachers and part of what they do (Lambert & Lines, 2000: p. 2). Therefore, assessment must be an integral part of instruction. Harmer (2001) thinks of teachers as assessors to fulfill the students' expectations through offering helpful feedback and correction on their performance. Assessment seems to drive teaching by forcing teachers to teach what is going to be assessed. Teaching involves assessment; that is, whenever a student responds to a question, offers a comment, or tries out a new word or structure, the teacher subconsciously makes an assessment of the student's performance (Brown, 2004). So when they are teaching, they are also assessing. A good teacher never ceases to assess students, whether those assessments are incidental or intended (2004: 4).

4. Assessment should serve learning.

Classroom assessment is an integral part of learning process as well. The ways in which learners are assessed and evaluated strongly affect the ways they study and learn. It is the process of finding out who the students are, what their abilities are, what they need to know, and how they perceive the learning will affect them (Sommer, 1989). In assessment, the learner is simply informed how well or badly he/she has performed (Ur, 1996). It can spur learners to set goals for themselves (Brown, 2001). Assessment and learning, as Feuerstein et al. (1980) stated, are seen as inextricably linked and not separate processes because of their mutually-influenced features (Williams & Burden, 2000). Learning by itself has no meaning without assessment and vice-versa. Thus, assessment places the needs of the students at the center of the teacher's planning (Penaflorida, 1998). The teacher designs the situation based on their assessment of the student's learning preference, interest, and needs. If learning is our central concern, then..., assessment should contribute to the learning process (Cameron, 2001).

5. Assessment should be curriculum-driven.

Classroom assessment should be the servant, not the master, of the curriculum (Lambert & Lines, 2000: 178). Assessment specialists view it as an integral part of the entire curriculum cycle. Therefore, decisions about how to assess students must be considered from the very beginning of curriculum design or course planning. L2 teachers see assessment as an activity which is integrated into the curriculum with the aim of improving learning, rather than a 'one-off' summative event (Cizek et al. 1995; cited in Brindley, 1998: 127). "The diagnostic or formative assessment is typically curriculum-driven so that assessment shadows the curriculum and provides feedback to students and teachers" (Spolsky, 1992: 38).

6. Assessment should be interactive.

Based on constructivist perspectives of second language learning, as students' cognitive activities are the core of teaching decisions, students should be proactive in selecting the content for assessment (Biggs & Tang, 1998). It provides a context for learning as meaning and purpose for learning and engages students in social interaction to develop oral and written language (O'Malley & Valdez Pierce, 1996). Feuerstein, considering assessment and learning as inextricably linked and not separate processes, maintained that "rather than viewing assessment as a process carried out by one person, such as a teacher, on another, a learner, it is seen as a two-way process involving interaction between both parties" (Williams & Burden, 2000: 42). Assessment, then, should be viewed as an interactive process that engages both teacher and student in monitoring the student's performance (Hancock, 1994).

7. Assessment should be student-centered.

Since learner-centered methods of instruction are principally concerned with language use and learner needs, students are encouraged to take more responsibility for their

own learning and to choose their own learning goals and projects. Therefore, in learner-centered assessment, they are actively involved in the process of assessment. Involving learners in aspects of classroom assessment minimizes learning anxiety and results in greater student motivation. Students in such classrooms "have input not only into what they learn, but also into how they will be assessed" (O'Malley & Valdez Pierce, 1996: 6). To make the classroom assessment possible, students should be encouraged to work on real and authentic tasks, interacting with others in class.

8. Assessment should be diagnostic.

Classroom assessment is diagnostic because teachers use it to find out learners' strengths and weaknesses during the in-progress class instruction. They also identify learning difficulties (Lambert & Lines, 2000: 109). If the purpose of assessment is to provide diagnostic feedback, then this feedback needs to be provided in a form – either verbal or written – that is for learners to understand and use (Brindley, 2003).

9. Assessment should be both bottom-up and top-down.

Learners' performance should be assessed in terms of both data-driven and conceptually-driven approaches because of the changing learning contexts. It depends on the nature of the knowledge the learners bring to the task. They don't seem to stick to one of these modes of processing in order to comprehend the language. These approaches are in constant interplay (Hedge, 2000) from the perspectives of interactive approaches to L2 learning. The learners, therefore, are in a continual shift of focus between them (2000: 109); that is, both schematic and systemic knowledge of the learners come into interaction in processing language. Thus, teachers should adapt their assessment procedures to the way learners process the language skills although the top-down approach is more prevalent in higher level skills.

10. Assessment should be exposed to learners.

Teachers are supposed to enlighten learners' accurate information about assessment. In other words, it should be transparent to learners. They must know when the assessments occur, what they cover in terms of skills and materials, how much the assessments are worth, and when they can get their results and the results are going to be used. They must also be aware of why they are assessed because they are part of the assessment process. Because the assessment is part of the learning process, it should be done *with* learners, not *to* them (Brindley, 2003). It is also important to provide an assessment schedule before the instruction begins.

11. Assessment should be non-judgmental.

In the classroom assessment, everything focuses on learning which results from a number of such factors as student needs, student motivation, teaching style, time on task, study intensity, background knowledge, course objectives, etc. So there is no praise or blame for a particular outcome of learning. Teachers should take no stance on determining who has done better and who has failed to perform well. Assessment should allow students to have reasonable opportunities to demonstrate their expertise without confronting barriers

12. Assessment should develop a mutual understanding.

Mutual understanding occurs when two people come to a similar feeling of reality. In second language learning, this understanding calls for a linguistic environment in which the teacher and students interact with each other based on the assessment objectives. Therefore, assessment has the ability to create a new world image by having the individuals share their thoughts helpful in learning process. When learning occurs, this is certainly as a result of common understanding between the teacher and students.

13. Assessment should lead to learner's autonomy.

Autonomy is a principle in which students come to a state of making their own decisions in language learning. They assume a maximum amount of responsibility for what they learn and how they learn it (Richards & Schmidt, 2002). Autonomous learning occurs when students have made a transition from teacher assessment to self-assessment. This requires that teachers encourage students to reflect on their own learning, to assess their own strengths and weaknesses, and to identify their own goals for learning (Genesee & Upshur, 1996). Teachers also need to help students develop their self-regulating and metacognitive strategies. Autonomy is a construct to be fostered in students, not taught, by teachers.

14. Assessment should involve reflective teaching.

Reflective teaching is an approach to L2 classroom instruction in which teachers are supposed to develop their understanding of teaching (quality) based on data/information obtained and collected through critical reflection on their teaching experiences. This information can be gathered through formative assessment (i.e., using different methods and tools such as class quizzes, questionnaires, surveys, field notes, feedback from peers, classroom ethnographies, observation notes, etc) and summative assessment (i.e., different types of achievement tests taken at the end of the term).

Conclusion

In this paper, I have addressed the fundamentals of assessment issues in L2 classroom settings. I should acknowledge here that assessment is a complex and progressing enterprise from both the theoretical and practical perspectives, and, thanks to the professional developments in education and especially teaching and testing, it calls for a greater demand for consideration than what has been indicated. But we should take it for granted that, in general, L2 teachers have not (been) developed in keeping with the research findings. One of the reasons is the complexities of an educational system shadowing all its areas, teaching, assessment, learning, planning, curriculum, etc, over which they are not able to have control adequately and equally. Another reason is that "the knowledge that teachers use in their practice, however, is more complicated than just knowing facts, using facts, and general conceptions of language and language learning" (Bartels, 2005: 1). More specifically, it is argued that the lack of success in assessment in L2 classrooms does not have to do with the teachers' ignorance of assessment issues only due to their unfamiliarity with measurement

techniques used for test analysis (Brindley, 1998). There are some other factors that are worth noting in this regard. It is a regret that L2 practitioners are starved of much of the latest findings in L2 assessments as well as the assessment literature, excluding the fact that most of them are not of use to the daily assessment practices of teachers. Also, we should not close our eyes to the fact that enough attention is not paid to the assessment in teacher education courses (Cizek et al. 1995; cited in Brindley, 1998: 127). In spite of all in the above, L2 teachers are believed to suffer a lack of having a firm knowledge of fundamentals of the classroom assessment. Accordingly, I tried to first provide the teachers with a succinct delineation of testing, assessment and evaluation in order for them to better understand what the concept 'assessment' really means in the current educational systems. Following this conceptualization, I called their attention to the understanding of under what theoretical foundation the currently-used assessment is justified.

I also illustrated the four issues effective in guaranteeing the assessment of students' actual ability in the classroom settings. It was argued that effective assessment requires L2 teachers to see if they assess what they are supposed to assess. Also, they need to ensure whether an assessment yields consistent results. Provision of equal learning opportunities to the students as well as determination of how much and to what degree an assessment may impact are the other two quality-oriented issues in effective classroom assessments that L2 teachers should take into careful account. Then, I explicated the purposes of assessment in two levels aiming for the internal and external decisions. One of the problems the L2 teachers encounter in actual teaching practices is the issue of assessment design. I outlined different steps so that teachers, especially the pre-service, in-service and novice ones, can follow them in actual practices. Since L2 teachers teach by principles, at least, from the research perspectives, there is no doubt they need to assess their students by principles, i.e., a set of beliefs and assumptions to make optimal decisions about their learning. I, therefore, pointed out fourteen principles helpful in administering the assessment of different types.

Since learning is the eternal issue of education, assessment is believed to be a means to ensuring the manifestation of that issue through the window of teaching process. To put it another way, it is the assessment, especially the internal enterprise, which changes the teaching behavior into the learning behavior. It seems to be a demanding process, which requires taking an account of many incidental and predictable factors. Therefore, practicing teaching on the basis of the fundamentals of L2 classroom assessment, as I explored in this paper, can, to a great extent, lighten the burden of the assessment and will result in a success in achieving both optimal learning and teaching. Putting these rudiments into practice also benefits the people outside the classroom setting, hoping that the effective administration of assessment of every type will yield positive impact in the educational community.

References

- Alderson, J. C. 1992. Guidelines for the evaluation of language education. In J. C. Alderson & A. Berreta (eds.), *Evaluating second language education*, (pp. 247-304). Cambridge: Cambridge University Press.
- Alderson, J. C. 2004. Forward. In L. Cheng, Y. Watanabe, & A. Curtis, (eds.), *Washback in language testing: Research contexts and methods*, (pp. ix-xii). Lawrence Erlbaum, Mahwah, NJ.
- Alderson, J. C., & Banerjee, J. 2001. Language testing and assessment (Part 1). *Language Teaching* 34, 213–36.
- Alderson, J. C., & Hamp-Lyons, L. 1996. TOEFL preparation courses: a study of washback. *Language Testing* 13(3), 280–97.
- Alderson, J. C., & Wall, D. 1993. Does washback exist? *Applied Linguistics*, 14(2), (pp. 155-129).
- Anderson, J. O., Muir, W., Bateson, D. J., Blackmore, D., & Rogers, W. T. 1990. *The impact of provincial examinations on education in British Columbia: General report*. Victoria: British Columbia Ministry of Education.
- Andrews, S. 2004. Washback and curriculum innovation. In L. Cheng, Y. Watanabe, & A. Curtis, (eds.), *Washback in language testing: Research contexts and methods*, (pp. 37-50). Lawrence Erlbaum, Mahwah, NJ.
- Bachman, L. F. 1990. *Fundamental considerations in language testing*. New York: Oxford University Press.
- Bachman, L. F. 2000. Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17, 1-42.
- Bachman, L. & Palmer, A. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bailey, K. M. 1996. Working for washback: a review of the washback concept in language testing. *Language Testing*, 13(3), 257–79.
- Bartels, N. 2005. Researching applied linguistics in language teacher education. In N. Bartels, (ed). *Applied linguistics and language teacher education*, (pp. 1-26). Springer. Springeronline.com.
- Berry, V. 1994. Current assessment issues and practices in Hong Kong: A preview. In D. Nunan, R. Berry, & V. Berry, (eds.), *Bringing about change in language education : Proceedings of International Language in Education Conference 1994* (pp. 31-34). Hong Kong: University of Hong Kong.
- Biggs, J. 1995. Assumptions underlying new approaches to educational assessment. *Curriculum Forum*, 4(2), 1-22.
- Biggs, J. 1996. *Testing: To educate or to select? Education in Hong Kong at the cross-roads*. Hong Kong: Hong Kong Educational Publishing.
- Biggs, J., & Tang, C. 1998. *Assessment by portfolio: constructing learning and designing teaching* . Hong Kong: Open University of Hong Kong Press.

Brindley, G. 1998. Language assessment and professional development. In C. Edler, et al. (eds.), *Experimenting with uncertainty: Essays in honor of Alan Davies*, (pp. 126-136). (Studies in language testing # 11). Cambridge: CUP.

Brindley, G. 2001. Assessment. In R. Carter, & D. Nunan, (eds.), *The Cambridge guide to teaching English to speakers of other languages*, (pp. 137-143). Cambridge: Cambridge University Press.

Brindley, G. 2003. Classroom-based assessment. In D. Nunan, (ed.), *Practical English Language Teaching*, (pp. 309-328). McGraw Hill

Brown, H. D. 2001. *Teaching by principles: An interactive approach to language pedagogy*. Englewood Cliffs, NJ: Prentice Hall Regents.

Brown, H. D. 2004. *Language assessment: Principles and classroom Practices*. US. Longman.

Brown, J. D. 2005. *Testing in language programs: A comprehensive guide to English Language assessment*. New edition. McGraw-Hill ESL/ELT.

Brown, J. D., & Hudson, T. 1998. The alternatives in language assessment. *TESOL Quarterly*, 17, 349-357.

Bryant, S. L., & Timmins, A. A. 2002. *Portfolio assessment: Instructional guide; using portfolio assessment to enhance student learning* (2nd ed.), The Hong Kong Institute of Education.

Buck, G. 1988. Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10, 12-42.

Cameron, L. 2001. *Teaching languages to young learners*. Cambridge: CUP.

Chappell, C. A., & Brindley, G. 2002. Assessment. In N. Schmitt, (ed.), *An introduction to applied linguistics*, (pp. 267-288). Oxford. OUP.

Cheng, L. 1998a. Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24(3), 279-301.

Cheng, L., & Curtis, A. 2004. Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis, (eds.), *Washback in language testing: Research contexts and methods*, (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum.

Cheng, L., Watanabe, Y., & Curtis, A. (eds.). 2004. *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum.

Cohen, A.D. 1994. *Assessing language ability in the classroom*. (2nd ed.). Boston, MA: Heinle and Heinle.

Coombe, C., Folse, K., & Hubley, N. 2007. *A practical guide to assessing English language learners*. The University of Michigan Press.

Cummins, J. P. 1980a. The cross-lingual dimensions of language proficiency: implications for bilingual education and the optimal age question. *TESOL Quarterly*, 14, 175-87.

Davies, A. 2003. Three heresies of language testing research. *Language Testing*, 20(4), 355-368

Educational Testing Service 2002. *ETS Standards for quality and fairness*. Princeton, NJ: Author.

Farhady, H. 2003. Classroom Assessment: A Plea for Change. Unpublished paper.

Frederiksen, J. R., & Collins, A. 1989. A system approach to educational testing. *Educational Researcher*, 18(9), 27-32.

Fulcher, G., & Davidson, F. 2007. *Language testing and assessment: An advanced resource book*. Routledge. Routledge Applied Linguistics.

Fullilove, 1992. The tail that wags. *Institute of Language in Educational Journal*, 9, 131-147.

Genesee, F. 2001. Evaluation. In R. Carter & D. Nunan, (eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 144-150). Cambridge: Cambridge University Press.

Genesee, F., & Hamayan, E. V. 1991. Classroom-based assessment. In E. V. Hamayan & J. S. Damico (eds.), *Limiting bias in the assessment of bilingual students* (pp. 212-239). Austin, TX: Pro-ED.

Genesee, F., & Upshur, J. A. 1996. *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.

Guba, E. G., & Lincoln, Y. S. 1989. *Fourth Generation Evaluation*. Newbury Park, CA: Sage Publications.

Hamp-Lyons, L. 1996. *Ethical test preparation practice: the case of the TOEFL*. Paper presented at the 18th Annual Language Testing Research Colloquium, Chicago, IL.

Hamp-Lyons, L. 1997. Washback, impact and validity: Ethical concerns. *Language Testing*, 14, 295-303.

- Hancock, C. 1994. : Eric Digest Clearinghouse on Languages and Linguistics. Washington DC.
- Harmer, J. 1998. *How to teach English*. Longman: Addison-Wesley.
- Harmer, J. 2001. *The practice of English language teaching*. (3rd ed.). London: Longman.
- Hasselgren, A. 2000. The assessment of the English ability of young learners in Norwegian schools: an innovative approach. *Language Testing*, 17(2), 261-277.
- Hedge, T. 2000. *Teaching and learning in the language classroom*. Oxford: OUP.
- Henning, G. 1987. *A guide to language testing: Development, evaluation, Reaearch*. Rowley, Massachusetts: Newbury House.
- Heyneman, 1987. Use of examination in developing countries: selection, research, and education sector management. *International Journal of Education Development*, 7, 251-263.
- Huerta-Macias, A. 2002. Alternative assessment: Responses to commonly-asked questions . In J. C. Richards, & W. A. Renandya, (eds.), *Methodology in language teaching: An anthology of current practice* (pp. 338-343). Cambridge: Cambridge University Press.
- Hughes, A. 1989. *Testing for language teachers*. Cambridge: CUP.
- Hughes, A. 2003. *Testing for language teachers*. Cambridge: CUP.
- Kimball, W. H., & Hanley, S. 1998. *Anatomy of a portfolio assessment system: using multiple sources of evidence for credentialing and professional development* . In N. Lyons, (ed.), *With portfolio in hand: validating the new teacher professionalism* (pp. 189-201). New York, NY: Teacher College Press.
- Klenowski, V. 1998. The use of portfolios for assessment in teacher education: A perspective from Hong Kong Asia Pacific. *Journal of Education*, 18(2), 74-86.
- Kunnan, A. 2005. Language assessment from a wider context. In E. Hinkel, (ed.), *Handbook of research in second language teaching and learning* (pp. 779-794). Mahwah, NJ: Erlbaum.
- Lambert, D., & Lines, D. 2000. *Understanding assessment: Purposes, perceptions, practice* . Routledge. Falmer.
- Lynch, B. K. 2001. The ethical potential of alternative language Assessment. In Catherine Elder et al. (eds.), *Experimenting with Uncertainty: Essays in honor of Alan Davies* (pp. 228-239). (Studies in language testing # 11). Cambridge: Cambridge University Press.
- McNamara, T. F. 1996. *Measuring second language performance*. London and New York: Longman.
- McNamara, T. F. 2000. *Language testing*. Oxford: Oxford University Press.
- Messick, S. 1989. Validity. In R. Linn (ed.), *Educational measurement*, Washington DC: Oryx and the American Council on Education.
- Messick, S. 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13-23.
- Messick, S. 1996. Validity and washback in language testing. *Language Testing*. 13, 241-56.
- Mousavi, S. A. 1999. *A dictionary of language testing*. (2nd ed). Tehran: Rahnama Publications.
- Nunan, D. 1999. *Second language teaching and learning*. Boston, MA: Heinle and Heinle.
- O'Malley, J. M., & Chamot, A. U. 1990. *Learning Strategies in second language acquisition* . New York: Cambridge University Press.
- O'Malley, J. M., & Pierce, L. V. 1996. *Authentic assessment for English language learners: Practical approaches for language teachers* . Reading, MA: Addison-Wesley.
- Penafiorida, A. H. 2002. Non-traditional forms of assessment and response to student writing: A step toward learner autonomy. In J. C. Richards, & W. A. Renandya, (eds.), *Methodology in language teaching: An anthology of current practice*, (pp. 344-353). Cambridge: Cambridge University Press.
- Pierce, 1992. Demystifying the TOEFL reading test. *TESOL Quarterly*, 17, 349-357.
- Popham, W. J. 1987. The merits of measurement-driven instruction. *Phi Delta Kappan*, 68, 679-682.
- Popham, W. J., Cruse, K. L., Rankin, S. C., Standifer, P. D., & Williams, P. L. 1985. measurement-driven instruction : It is on the road. *Phi Delta Kappan*, 66, 628-634.
- Rea-Dickins, P. 2001. Mirror, mirror on the wall: identifying processes of classroom Assessment. *Language Testing*, 18(4), 429-462.
- Richards, J. C., & Renandya, W. A. (eds.). 2002. *Methodology in language Teaching: An anthology of current practice* . Cambridge: Cambridge University Press.
- Richards, J.C., & Rodgers, T. 2001. *Approaches and methods in language teaching*.

(2nd ed.). Cambridge: Cambridge University Press.

Richards, J. C., & Schmidt, R. 2002. *Longman dictionary of language teaching and applied linguistics*. (3rd ed.). Pearson Education. Limited.

Shepard, L. 1991. Psychometricians' beliefs about learning. *Educational Researcher*, 20(6), 2–16.

_____. 1993. Evaluating test validity. *Review of Research in Education*, 19, 405–50.

Shohamy, E. 1993a. The exercise of power and control in the rhetorics of testing. In A. Huhta, K. Sajavaara, & S. Takala, (eds), *Language testing: new openings*.

Jyvaskyla: University of Jyvaskyla, (pp. 23–38).

_____. 1994. The validity of direct versus semi-direct oral tests. *Language Testing* 11 (2), 99–123.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. 1997. Test impact revisited: washback effect over time. *Language Testing* 13(3), 298–317.

Smith, M. L. (1991b). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.

Sommer, R. F. 1989. *Teaching writing to adults*. San Francisco: Jossey-Bass.

Spolsky, B. 1992. Diagnostic testing revisited. In E. Shohamy, & R. A. Walton, (eds.), *Language assessment and feedback: Testing and other strategies*, (pp. 29-39). National Foreign Language Center. Dubuqu, IA: Kendall/Hunt

Publishing Co.

_____. 1994. The examination-classroom backwash cycle: some historical cases. In D. Nunan, R. Berry, & V. Berry (eds.), *Bringing about change in language education : Proceedings of International Language in Education Conference 1994* (pp. 55-66). Hong Kong: University of Hong Kong.

_____. 1995. *Measured Words: The Development of Objective Language Testing*. Oxford, Oxford University Press.

Stern, H. H. 1983. *Fundamental concepts of language teaching*. Oxford: Oxford University Press.

Stiggins, R. 1994. *Student-centered classroom assessment*. NY: Macmillan.

Tierney, R. L., Carter, M. A., & Desai, L. E. 1991. *Portfolio Assessment in Reading-Writing Classroom*. Norwood, Mass: Christopher-Gordon.

Ur, p. 1996. *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.

Valdez Pierce, L., & Gottlieb, M. 1994. Portfolios: Matching purpose to practice. Proceedings. Baltimore, Md.

Valencia, S., Hiebert, E., & Affleberg, P. 1994. *Authentic reading assessment: practices and possibilities*. Newark, DE: International Reading Association.

Valette, R. M. 1994. Teaching, testing, and assessment: conceptualizing the relationship. In C. Hancock, (ed.), *Teaching, testing, and assessment: making the connection* (pp. 1-42). Northeast Conference on the Teaching of Foreign Languages.

Wall, D. 1996. Introducing new tests into traditional systems: insights from general education and from innovation theory. *Language Testing* 13(3), 334–54.

_____. 1997. Impact and washback in language testing. In C. Clapham (ed.). *The Kluwer encyclopedia of language in education: Vol. 7. Testing and assessment* (pp. 291–302). Dordrecht, Netherlands: Kluwer.

_____. 2000. The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System*, 28, 499–509.

Wall, D., Alderson, J. C. 1993. Examining washback: the Sri Lankan impact study. *Language Testing*, 10(1), 41-69.

Watanabe, Y. 1992. Washback effects of college entrance examinations on language learning strategies. *JACET Bulletin*, 23, 175–194.

_____. 1996a. Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13, 318–333.

_____. 1996b. Investigating washback in Japanese EFL classrooms: Problems and methodology. *Australian Review of Applied Linguistics*, 13, 208–239.

_____. 2004. Teacher factors mediating washback. In L. Cheng, Y. Watanabe, & A. Curtis, (eds.), *Washback in language testing: Research contexts and methods*, (pp. 129-146). Mahwah, NJ: Lawrence Erlbaum.

Widdowson, H. G. 1979. *Explorations in applied linguistics 2*. Oxford: Oxford University Press.

Wiggins, G. (1989a). Teaching to the (authentic) test. *Educational Leadership*, 48(7), (pp. 41-47).

_____. 1991. Standards, not standardization: Evoking quality student work. *Educational Leadership*, 48(5), (pp.18-25).

_____ 1992. _____. Educational Leadership, (8), (pp. 26-33). In J. M. O'Malley, & L. V. Pierce, *Authentic assessment for English language learners* . Reading, MA: Addison-Wesley.

_____ 1993. *Assessing student performance*. San Francisco, CA: Jossey-Bass.

Williams, M., & Burden, R. L. 1997. *Psychology for language teachers: A social constructivist approach* . Cambridge University Press.

Wolf, K. P. 1993. From informal to informed assessment: Recognizing the role of the classroom teacher. *Journal of Reading*, 36(7), 518-523.

Worthen, B. R. 1993. Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan*, 74(6), 444-454.