

Theoretical Framework for Educational Assessment: A Synoptic Review

Abdallah Ghaicha*

Faculty of Letters and Humanities, Ibn Zohr University, PO box 29/S, Hay Addakhla, Agadir, Morocco

Abstract

At this age of accountability, it is acknowledged that assessment is a powerful lever that can either boost or undermine students learning. Hitherto, much of the regular institutional and instructional practices show that assessments remain inhibitory or void rather than constructive as these lack the assessment formative aspect. This denotes that assessment is either not well understood or not done in a principled educational framework across all educational levels. This is due to some inherited dysfunction of the past, which calls for the urgency of a moratorium. The current article attempts to enhance the EFL practitioners' educational assessment practice as a form of expanding awareness. It attempts to establish a founded theoretical framework for the main concerns that might have been troubling novice and professional EFL practitioners with regard to understanding the working mechanisms of such a perplexing task that has long been delegated to them. Particular aspects of the current review target the definitions of the concept of assessment, the value, functions and purposes of assessment, levels where assessment occur, assessment research literature synopsis, and classroom assessment research (CAR). CAR presents detailed knowledge about the potency of assessment, research on classroom assessment practices, research on alternative assessment, research on formative assessment, and finally quality control criteria for effective classroom assessment. The article culminates in the pedagogical potency of formative assessment and some of its classroom procedural applications inspired from research into formative assessment and how that might contribute to the enhancement of pedagogic practice.

Keywords: educational assessment – accountability – alternative assessment – formative assessment – assessment practice

1. Introduction

To carry out research on educational assessment (henceforth EA) in higher education, it will be supportive to have some theoretical considerations at hand. Central to EA are the dichotomous concepts like testing, evaluation, assessment, and examination. A rich body of literature on CA (practice) research has been reviewed in order to establish a founded theoretical framework for the main concerns that might have been troubling novice and professional EFL practitioners with regard to understanding the working mechanisms of such a perplexing task that has long been delegated to them. Particular aspects of the current review target the definitions of the concept of assessment, the value, functions and purposes of assessment, levels where assessment occur, assessment research literature synopsis, and classroom assessment research (CAR). CAR presents detailed knowledge about the potency of assessment, research on classroom assessment practices, research on alternative assessment, research on formative assessment, and finally quality control criteria for effective classroom assessment. Stated below are the questions that were chosen to drive the theoretical search:

1. What does the EA literature say about assessment of student learning?
2. What makes a classroom assessment effective?
3. What quality control criteria that language teachers need to assure to make an assessment of high quality?

It is worth noting that even though the body of the literature that was reviewed is not directly related to the practice of EFL testing in higher education as most of the (research) publications available to the researcher target the issue of assessment in primary and secondary levels, I believe that the same principles could apply to EFL testing at higher levels of education.

2. Definition of the concept of Assessment

A number of specialized books, journals, seminal articles, conference papers, currently defended theses and dissertations were scrutinized in an attempt to find a comprehensive definition of the concept of assessment. These references and other day-to-day classroom practices exhibit a number of functions, forms, tools and techniques available to teachers, as classroom assessors, as well as numerous terms, phrases, concepts and descriptions of assessment. Most of them seem to be confusing to people who are unfamiliar with the jargon of EA. Norm, criterion, formative, summative, traditional, standardized, authentic, alternative, performance, balanced, etc. have each added to the knowledge base of assessment, but bewildered both novice researchers and authorities on the subject. To demystify these concepts and other pertinent concepts and issues previously mentioned in this section, a brief overview of these concepts is presented below.

To begin with, terms like evaluation, measurement and testing have been closely associated with and related to assessment. They are even sometimes used interchangeably as means used to gather information on

student learning. A layman or an average observer may think that they have the same meaning, but there are distinct differences. According to Mundrake (2000), "Assessment, testing, and evaluation are terms used to describe the outcomes of the educational process" (p. 45). Mundrake (2000) further notes "Assessment is the term currently used to describe all aspects of evaluation and testing" (p. 45). So, what distinguishes one from the other?

According to Bachman (2004), "The term '*assessment*' is commonly used with a variety of different meanings. Indeed, the term has come to be used so widely in many different ways in the field of language testing and educational measurement that there seems to be no consensus on what precisely it means" (p.6). Brown (2004) defined assessment as "any act of interpreting information about student performance, collected through any of a multitude of a means or practices" (p. 304).

Furthermore, a number of other terms are frequently used more or less synonymously to refer to assessment. For the purpose of this thesis, assessment is operationally defined as a part of the educational process where [faculty] instructors appraise students achievements by collecting, measuring, analyzing, synthesizing and interpreting relevant information about a particular object of interest in their performance under controlled conditions in relation to curricula objectives set for their levels, and according to the procedures that are systematic and substantively grounded. It requires assigning students' performances numerical descriptions of the extent to which they possess specific characteristics or traits measured according to specific standards, or criteria serving as a source of evidence of many aspects of an individual student's knowledge, understanding, skills and/ or abilities. Such information can be elicited through any of a multitude of means or practices and other measures recommended by the educational system- involving activities of teachers, students, a written test paper, an interview schedule, a measurement task using equipment, a class quiz. It should serve as a form of communicating feedback both to students' learning and teachers' teaching. "In the classroom, assessment considers students' performances on tasks in a variety of settings and contexts". It is the most general of the terms that describe how teachers gather and use information. This process usually involves a range of different qualitative and quantitative techniques. For example, the language ability of learners can be assessed using standardized tests (pen/ pencil and paper exam, oral exams, portfolios, and practical exercises, etc.

Evaluation refers to the process of arriving at judgments about abstract entities such as programs, curricula, organizations, institutions and individuals. For example, systemic evaluations are conducted to ascertain how well an education system is functioning. In most education contexts, assessment is a vital component of any evaluation. It is the process of judging the quality of content and programs offered to a group of students. Teachers usually assess students and use this assessment information to judge the quality of student learning for summative or formative purposes. High quality evaluations do not necessarily require the use of pen-and-pencil tests or examinations. Neither do they require the use of complex measurement approaches. Of course, evaluations may use information from tests and measurement. It is an open question whether teacher-made evaluations are improved by using any or both of tests and measurements.

Another term that is often associated with assessment is *measurement*. It is the process by which a quantified value, usually numerical, is assigned to the attributes or dimensions related to students' performance while measuring ability or aptitude in such a way that the students quality of performance is preserved (Bachman, 2004; Nitko, 1996; Airasian, 1994). Gallagher (1998) is even more specific when she says "measurement is the process of quantifying the degree to which someone or something possesses a characteristic, quality, or feature" (p.3) It can be done by counting how many correct responses a student gives in relation to the total, or by assigning a percentage, or by assigning a student a numerical score. Yet, not all assessment requires the measurement of students and assigning marks or scores to them.

Comparatively, *testing (or examining)* is the process of administering a test to elicit and measure a certain behavior (concept) from which one can make inferences about certain characteristics of an individual, usually under standardized conditions. For example, tests are used to measure how much a student has learned in a given course or subject by means of more or less formal, systematic methods of assessment used to determine a student's knowledge with regard to a predetermined content. Most often, these methods require the use of paper-and-pencil instruments designed to elicit some definite behavior, knowledge, or skill from the test taker. Linn and Gronlund (1995) describe the test as "a type of assessment that typically consists of a set of questions administered during a fixed period of time under reasonably comparable conditions for all students" (p.5). Sometimes the results of assessing students are reported on a numerical scale reflecting quality of learning through a quantitative score or mark. Higher grades reflect higher levels of learning or competence; whereas lower grades reflect a deficiency or incompetence related to the target content.

3. Understanding the value, Functions and Purposes of Assessment

Educational institutions worldwide, across all educational levels, are involved, to some extent in the development and implementation of some kind of academic assessment. This involvement may be mandatory, voluntary, or a combination of both.

The value of EA lies in its judgmental and instructive roles for both authorities and individuals. From the

perspective of authorities, assessments allow people in charge of assessment to get instructive insights about three critical functions: Selecting, Monitoring, and Holding Accountable. Assessment results, along with other measurement data (such as those obtained through periodic surveys), are valuable tools for educational institutions. They assist in evaluating the effectiveness of institutional practices by tracking the functioning of different components of the assessment system (generally referred to as national assessments), sometimes holding the individuals responsible for those components accountable, contributing to decision making about the functioning of departments, programs and curricula, and providing potential measures to be taken to improve all the cornerstones of an educational system. It can also play a role in planning . . . , in budgeting . . . , and in faculty rewards. . .” (Murphy & Harrold, 1997). It also allows departments, or division heads to evaluate the effectiveness of entire programs and allows faculty to determine what, and how well, students are learning and how effective are both their instructional and assessment practices are to their students and the accountability measures set by their educational system. From the perspective of individuals, good assessments assist teachers in evaluating the effectiveness of their teaching, providing them with a framework to fine tune teaching methods (Zeliff, 2000). It also monitors student progress and achievement, determines the performance levels of individual students and teachers, and controls program evaluation and curriculum review in an effort to improve instruction and teacher effectiveness.

Classroom assessments are not run in void. They are governed by the purposes, uses and functions to which they are put. Some language testers have argued that the most important consideration in developing, adopting, or adapting educational tests is the uses to which test-based information will be put (e.g., Bachman & Palmer 1996; Brown 1996), and it is becoming increasingly apparent that there are many possible uses (indeed, many required ones) for assessments in contemporary college and university foreign language (FL) assessments, as in all of education.

Various sources on FL EA, for example, have regularly paid perfunctory homage to the notion that assessment might need to meet distinct purposes, although in doing so, they have typically produced disparate lists of exactly what those purposes should be. For example, Lado (1961) cited achievement, diagnostic, and aptitude uses for language tests while Larson and Jones (1984) distinguished between proficiency and achievement tests. Three main assessment roles in college FL education as placement, achievement, and proficiency testing have too been described in the literature.

Many FL educators have drawn a basic distinction between formative and summative assessments (e.g., Swain, 1984; Omaggio, 1986); although for some this is simply a temporal distinction rather than the identification of different ways of using assessment information (e.g., ACTFL 2002). Others have discussed a much wider variety of uses for assessment. For instance, Finnochiaro and Sako (1983) provided over fifteen distinct answers to the question “Why do we Test?” and then categorized these according to three overarching uses of language test results: a) student measurement; b) instructional evaluation; and c) curriculum evaluation. Nitko (1996), as displayed in Figure 1 below, shows how these purposes are related to EAs. The branch of the figure associated with decisions about students shows several categories of decisions such as decisions about managing their instruction, placing them into special educational programs, and selecting them for further educational opportunities. The figure elaborates the managing instruction decision category to identify more specific decisions for which teachers need assessments, including planning instruction, placing students into learning sequences, and assigning final marks or grades.

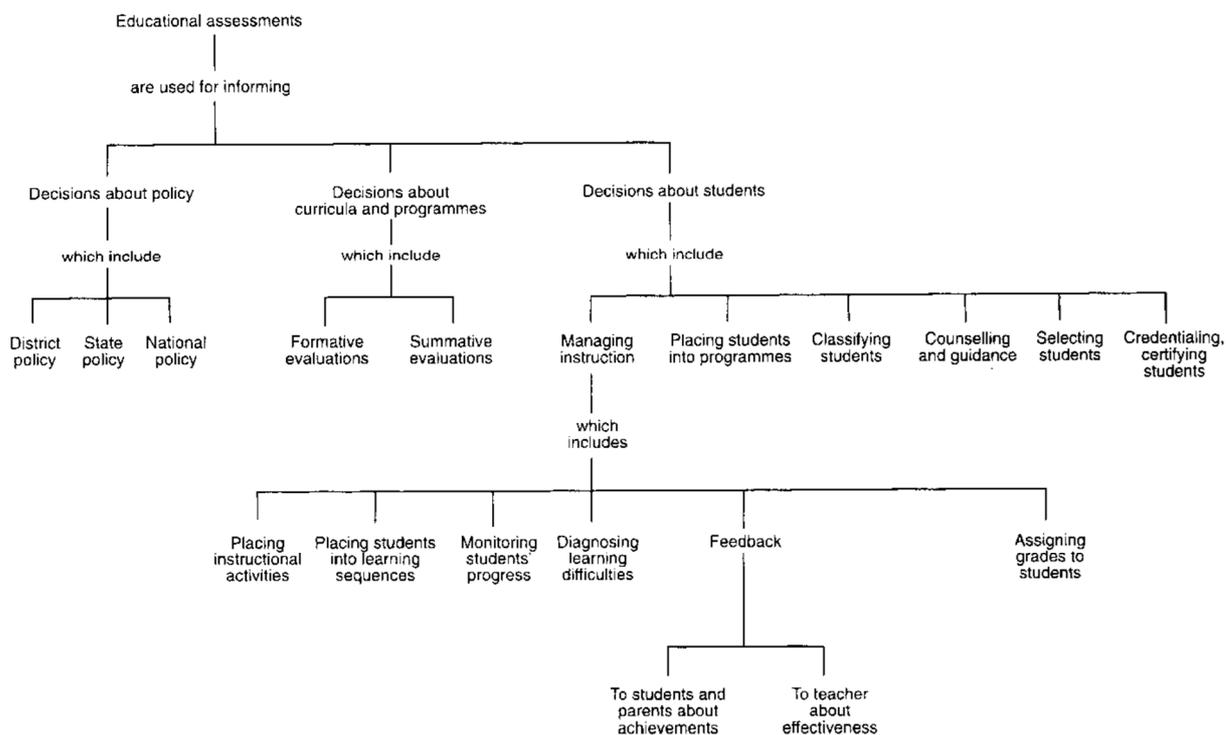


FIG. 1. Examples of types of educational decisions for which assessments may be used. Source: Nitko (1996).

Each of these examples, and similar ones like them within FL education, has indicated the potential diversity in applications for FL assessments, ranging from very generic distinctions between types of tests to much more particular delineations of their specific intended uses. However, as these sources have each gone on to describe the characteristics that constitute good assessment practice they have generally failed to consider that this diversity in intended uses for assessment might call for a parallel diversity in the form of distinct designs that emphasize unique qualities of assessment most appropriate and effective for meeting these purposes. Instead, they have emphasized the need for tests to be designed according to traditional principles of good measurement-objectivity, reliability, fairness, etc., in short, the validity characteristics that make a test (any test) a good measure of what it was intended to measure.

Instructors assess for various reasons (Airasian, 1994; Linn & Gronlund, 1995; Pellerino, Chudowsky & Glaser, 2001). School and public policies require instructors to conduct a type of assessment designed to meet the purpose of gathering more information and making decisions about their students. These assessments may assist instructors in making judgments about student behavior and academic performance, as well as in diagnosing their strengths and weaknesses, which in turn may result in adjustments within the classroom in the form of remedial work or referral for outside assistance. Phye (1997) summarizes three purposes of assessment: “(a) discovering and documenting students’ strengths and weaknesses, b) planning and enhancing instruction, and c) evaluating progress and making decisions about students” (p. 10). Kane, Khattri, Reeve, and Adamson, (1997) identify the four purposes of assessment as: influencing and informing instruction and curriculum, monitoring student progress, holding teachers and schools accountable, and certifying student achievement. Assessment can also serve to provide feedback to students by measuring their progress, giving them an idea of their degree of (non)mastery of the content taught to them in relation to others or to a norm or a standard. Instructors can use assessment to place students in a group for behavioral, social and/or instructional purposes. This can be achieved through conducting either or both summative or formative assessment.

There is a clear gap between prescribed notions of good measurement in FL education and the ways in which these good measures might or might not be applied in resolving the actual problems of unique measurement uses, purposes, and intended consequences in FL settings. Meanwhile, whether or not the characteristics and qualities of language measures are aligned with their purposes and uses, assessments are regularly applied within all levels of FL settings (a) for making real decisions about students, (b) for informing instruction and learning, and (c) for meeting increasing demands for accountability and program improvement from within or outside institutional walls.

4. Levels Where Assessment Occurs

To complement the categorization of the different roles of assessment, a brief overview of the different types¹⁰ of assessments that are typically employed by educational systems worldwide and the levels at which they occur is presented below.

For a comprehensive assessment effort to function, governmental officials must gain a thorough understanding of its overall effectiveness as well as the effectiveness of individual programs accredited and made parts of the policy and curriculum being offered to national or local educational institutions. Therefore, assessments are hierarchically set in a top-down model, each feeding the other: national assessments and institutional or school (classroom) assessment. The latter will be explored in detail below as it is the main concern of this dissertation.

National assessments are studies focused on generating specific information that policymakers need to evaluate various aspects of the educational system: accountability of the teaching staff, efficiency of programs and curricula, and the accountability of both with regard to student academic achievement. The results can be used for accountability purposes, to make resource allocation decisions, and even to heighten public awareness of educational issues. These assessments may be administered to an entire cohort (census testing) or to a statistically chosen group (sample testing) and may also include background questionnaires for different participants (learners, teachers, administrators) to provide a meaningful context for interpreting test results. The utility of the data generated depends on the quality and relevance of the assessment, the thoroughness of the associated fieldwork, as well as the expertise of those charged with the analysis, interpretation, reporting, and dissemination of results. To get beneficial results, national assessments must use other (types of) independent assessments that can be taken locally at the institutional level. The purpose of such an assessment and the credentials of the board or group responsible for it should guide these assessments.

Comparatively, the most common type of assessment is classroom-based assessment, which is institution-based. The remaining sections of this chapter will provide a theoretical overview of classroom assessment research. Next is a synopsis of assessment research literature.

5. Assessment Research Literature: Synopsis

The constellation of assessment literature, research and empirical evidence on classroom assessment over the world has provided myriad points of illumination. There are historical perspectives that run from the back-to-basics movement of the 1970s, through the minimum competency movement of the 1980s and the growth of standardized testing in the 1990s, especially in USA, to the reauthorization of the Elementary and Secondary Education Act in USA in 2001 and the move to the introduction of alternative modes of assessment (Marzano, Pickering, & McTighe, 1993; Phye, 1997; Linn, 2001).

Indeed, assessment plays an integral role in contemporary educational practice, and its use is ubiquitous among educators across all formal education contexts. Policymakers hold a remarkable view of the value of standardized testing in regulating education and holding schools accountable to the learners and the system (McDonnell, 1994). Psychometricians offer statistical strategies for enhancing validity and reliability of tests (Clare, 2000; Linn & Baker, 2001). Teachers and students also devote a large proportion of their class time to assessment activities (Stiggins & Conklin, 1992), and most of what students, teachers, and others (e. g., stakeholders, parents, researchers) know about what students are learning, what they are capable to do, and what decisions are made about them comes from the use of assessments (Brookhart, 2003; National Research Council (NRC), 2001). Increasingly, practitioners in all educational levels, including teachers, administrators, policymakers, as well as system evaluators, are expected to understand the principles of assessment (and be certified in them via assessment) and to engage in sound assessment practices within and beyond the classroom (e.g., Elliot, 2003; NCATE, 2002; Schafer & Lissitz, 1987; Stiggins 1999; Wise, 1993).

Approaches to curriculum and instruction are more apt to be integrally linked with, or even driven by, assessment practices and the forms that they take (e.g., Angelo & Cross, 1993; Huba & Freed, 2000; Wiggins & McTighe, 1998). Furthermore, current educational policy at all levels has been gripped by a veritable frenzy to hold students, teachers, programs, schools, and institutions accountable to the public through assessments (e.g.,

¹⁰ *Public examinations* is another type of assessment that can fulfill one or more of the following roles: selecting learners for admission to next or higher levels of education or training centers for professional careers, credentialing learners for the world of work, and/or providing data for holding school staff accountable for their performance. While such examinations are an important component of every nation's education system, they are particularly critical in (under)developing countries, where the number of candidates for advancement is usually many times greater than the number of places available. In many countries, these are standardized multiple choice examinations, while in others they comprise various forms of performance assessment (sometimes in conjunction with multiple choice components). Typically, they are designed, developed, and administered centrally with an almost exclusive focus on academic subjects. There is meager feedback to the school except the scores and/or pass rate, and, as a result, they offer little utility for school improvement programs beyond an exhortation to do better next time. Moreover, as we have already noted, public examination systems often have negative consequences for the general quality of education.

Shavelson & Huang, 2003; Brown, 2004), and EA professionals have devoted extensive energies to the development of assessments based on professional, and national standards of student learning across all content areas and disciplines (e.g., Cizek, 2001, 2003; Phelps, 1998; Popham, 1999).

The literature is also replete with guidance on how to write a good test (e.g., Airasian, 1994; Alderson, 1988, 1990a, 1990b, 2000; Alderson & Banergee, 2001; Alderson *et al.*, 1991; Alderson *et al.*, 1995; Bachman, 1990; Bachman, 1985; Bachman, 1991; Bachman & Palmer, 1996; Gallagher, 1998; Linn & Gronlund, 1995; Banergee & Luoma, 1997). The balance among these views of assessment is manifested in the literature on authenticity in assessment (Wiggins, 1990; Aschbacher, 1993; Shepard, Flexer, Hiebert, Marion, Mayfield, & Weston, 1995; McTighe, 1996; Eisner, 1999). Phye (1997) contends that combining all these factors has resulted in "classroom assessment being one of the most hotly debated topics in educational circles" (p. 33).

Along similar lines, even in the once 'off-limits' arena of higher education (including graduate education, see Haworth 1996), faculty are being asked to engage in the structured and well-informed and principled assessment of student learning, not only for the purposes of degree-program and institutional accreditation (e.g., Chun, 2002; Maki, 2002; Peterson & Einarson, 2001), but also as a means for understanding and improving student learning (hence formatively improving instruction), and for revising curricular and instructional practices (e.g., Cross, 1999; Lopez, 1998; Suskie 2000).

Despite this ubiquity, or perhaps because of it, what constitutes a 'good' or 'appropriate' assessment in education has proven to be a highly contentious question, the answers to which have ranged considerably depending on the purposes, uses, users, and contexts for assessment (see section I. 7 below for a detailed section about effective classroom assessment). Indeed, even arriving at a definition of assessment invites disagreement, with some arguing sharp technical distinctions among terms like assessment, measurement, evaluation and testing (e.g., Embretson & Hershberger, 1999), and others like Popham (2000), contending that each of these terms can be defined identically, at least within education, as "a process by which educators use students' responses to specially created or naturally occurring stimuli in order to make inferences about students' knowledge, skills, or affective status" (p. 3).

As definitions have differed, so too have prioritized qualities of EA. For example, some have argued that large-scale, norm-referenced testing of students based on national standards of achievement will ensure the accountability of teachers and schools, resulting in positive consequences for student learning outcomes (e.g., Cizek, 2001, 2003; Mehrens, 1998). However, others have countered that this focus on accountability, and the concomitant demand for highly discriminating instruments based on traditional testing formats, is inappropriate for evaluating the quality of schools and teachers, and that it causes a reductionist approach to curriculum and the denigration of instruction and learning (e.g., Bryk, 1998; Popham, 1999, 2003b). In a similar vein, so-called 'alternative' assessments, including in particular performance and portfolio assessment, have been advocated for use in both classroom- and curriculum-based assessment as well as for high-stakes achievement testing (e.g., Aschbacher, 1991; Wiggins, 1989, 1993b; Wolf, Bixby, Glenn, & Gardner, 1991). By assessing authentic, complex performances and samples of student work, it has been argued, teachers and students will focus on valued learning outcomes as opposed to figuring out how to score well on test items that have little to do with such outcomes. In response, others have emphasized perceived problems with reliability, domain sampling, population biases, and related concerns in suggesting that such 'alternative' assessments may not provide the most appropriate alternatives (Cizek, 1991; Eisner, 1999; Haertel, 1999; Mehrens, 1992). Still others have promoted distinct qualities for EAs, including, for example, a focus on the feedback potential of classroom based assessments (e.g., Angelo & Cross, 1993), a prioritization of educative¹¹ over auditing properties of assessments designed to improve student learning (e.g., Wiggins, 1998), and the clarity of objectives and relevance for instructional decision making of criterion-referenced and curriculum-based assessments (e.g., Glaser 1994; Nitko, 1995, 2001; Popham, 1994).

With the increasing emphasis on classroom assessment, there is a growing movement towards balanced assessment systems. Local (within institution) assessment systems can provide more detailed information about individual students that can be used to improve instruction (Rabinowitz, 2001; Cutlip, 2003). These locally developed systems have many supporters and recommendations for their design (Coldarci, 2000; Rabinowitz & Ananda, 2001; Stiggins, 2002b; Cutlip, 2003) but relatively little research on validity and reliability of such assessment systems. In effect, there is little data to show whether teachers are aware of specific components, are applying them appropriately in the classroom, and are meeting the recommended standards (Cromey & Hanson, 2000). This crossroads, where standardized testing and classroom assessment blend with knowledge of other disciplines to contribute to the development of assessment systems, raises issues pertinent to regular classroom

¹¹ **Educative assessment:** a term used by McMillan (2000) to expand on previous descriptors of quality assessment to also include the ability to distinguish between the measurements (scores) and evaluation (interpretation of scores). He uses Grant Wiggins' term "educative assessment" to explain that assessment influences student learning, engagement, and motivation. Not only does assessment inform instruction, but also enhances it. He encourages the use of multiple measures that are fair and ethical as well as efficient and feasible.

practices of assessment.

In higher education, literature reviews have documented that an increasing number of colleges and universities have engaged in some form of student assessment activity since the beginning of the 1980s (El-Khawas, 1988, 1990, 1995). Administrators and faculty have invested considerable time and effort in promoting, supporting, and implementing these assessment efforts. However, concerns have been raised that assessment activities are difficult to mount successfully (Gray & Banta, 1997; Ewell, 1988a) and seldom produce discernible impacts on student or institutional performance (Astin, 1991; Ratcliff et al., 1995).

Student assessment, particularly scholarship taking an institutional perspective, is an emerging arena of study in higher education. Scholars and practitioners have produced many books, seminal articles, monographs, theses and dissertations that offer prescriptive guidelines for how institutions might best approach and support their student assessment efforts (cf. Rossman & El-Khawas, 1987; Sims 1992; Banta & Associates, 1993; Banta, Lund, Black, & Oblander, 1996). These efforts remain theoretically prescriptive. Less available is nationally representative empirical evidence concerning how institutions have conducted student assessment and to what effect as well as how principled are the classroom assessments that faculty instructors develop or use to assess students' achievements in their classes. Studies have collected descriptive data regarding the content and methods of institutions' assessment approaches (Ory & Parker, 1989; Cowart, 1990; Steele & Lutz, 1995). But there has been little systematic examination of organizational and administrative patterns at the institutional level developed to support student assessment efforts (Johnson, Prus, Andersen, & El-Khawas, 1991; Ewell, 1997) or how institutions have used and been affected by assessment information (Ewell 1988b; Banta *et al.* 1996; Gray & Banta, 1997).

Some theorists believe that teachers' assessment literacy is limited (Plake & Impara, 1997). They receive little training in classroom assessment (Stiggins & Conklin, 1992; Stiggins, 2001). Phye (1997) claims that a "teacher's assessment grading practices are highly variable" (p. 29). Peterson and Einarson (2001) compensated for this dearth by extending current understanding of how postsecondary institutions have approached, supported, and promoted undergraduate student assessment, and the institutional uses and impacts that have been realized from these assessment efforts. They examined the congruence between institutional approaches to student assessment found in the prescriptive literature and actual institutional practices. They focused on five conceptual domains: (1) the relationship of institutional context to student assessment; (2) institutional approaches to student assessment; (3) organizational and administrative support for student assessment; (4) assessment management policies and practices; and (5) institutional uses and impacts of student assessment information. They offered useful recommendations for institutions embarking on student assessment and empirical literature providing evidence of institutional practices and consequences with respect to student assessment. Yet, the current literature enlightens only portions of the picture, raising more questions than answers about teachers' utilization of assessment in the classroom. The research, as is shown in this synoptic review, is particularly lacking at the higher education level. Practical insights will be drawn from the existing classroom assessment research in other levels of education.

6. Classroom Assessment Research

6.1. The Potency of Classroom Assessment

Classroom assessments serve educational practice through the numerous functions they fulfill. These assessments are usually devised and administered by class instructors, although some are the work of the department heads, boards, or groups of teachers or other instructional staff. Typically, they should be aligned with the delivered curriculum and may employ a broader array of media- addressing a greater range of topics than is the case of standardized assessments- to assess (a) *course-related knowledge and skills*, for instance by employing pre-conception checks, minute papers, word journals and concept maps, written tests, oral presentations, portfolios, projects, (b) *learner attitudes*, values and self awareness, using approaches such as classroom opinion polls, self-confidence surveys, skills-checklists, and self-assessments of learning styles, and (c) *learner reactions to instruction*, including the use of teacher designed feedback forms, quality circles and assignment assessments.

Classroom assessments have a decided advantage over centralized assessments in that the results are immediately available to the teacher (and, presumably, the learners) and can influence the course of instruction for both. While these assessments can play an important role in promotion to the next grade, they are rarely used for high-stakes decisions such as admission to the next level of the education system. Such a type of assessment can have a great potential in accelerating learning for all learners. Other decisions, like program assessment, need be made on a shared basis at the national and institutional level. The prime concern is the evaluation of an overall academic program. Assessment can have a tremendous effect on the curriculum being offered. It can help academic units in two ways: (1) by affirming that things are going well in terms of the curriculum and courses offered, and (2) by identifying things that are not going so well. Thus, individual courses and/or programs can be added, removed, or modified. Individual course structures or teaching methods can be evaluated and changed if necessary. The appropriateness of course requirements, prerequisites, and sequencing can be determined.

Faculty members are critical agents in assessing and analyzing student learning and can provide such information to others in their respective academic programs/units as well as to the institution as a whole. Wolvoord, Bardes, and Denton (1998) indicate that teachers have power when it comes to generating information about classroom learning. Institutions definitely are interested in such information.

As the foregoing exposition should make clear, assessment has the potential to contribute to one or more aspects of the educational sphere, depending on the type, purpose, level and context in which it is employed. School-based assessments can enhance efficiency by helping to target the efforts of both learners and teachers to the extent that they are able to use the information appropriately; the quality of the learning is improved. National assessments are said to be in the service of decisive policies as they are running evaluations on a detailed, but global, basis attempting to get deeper into accountability concerns.

Classroom assessment has, however, been influenced by historical forces, standardized testing, authentic assessment, reform movements, and a myriad of other factors. It is important to understand how these factors have contributed to the present status of classroom assessment and how they are driving the future of assessment. This information then becomes the basis for understanding the current status of classroom assessment. Detailed information about classroom assessment research will be provided below, but with limited focus on classroom assessment practice, alternative (authentic or performance) assessment research, formative assessment and its potential for instruction, and quality control standards for effective assessment.

6. 2. Research on Classroom Assessment Practices

The research on classroom assessment practices has evolved in tandem with the ongoing assessment reforms that the EA research witnessed in the recent years. Prior to the mid- 1980's the assessment literature "focused almost exclusively on large-scale standardized testing" (McMillan & Workman, 1998). These early studies noted that teachers were using selected response items most frequently and most questions sampled factual and conceptual knowledge (Airasian, 1984; Airasian, Kellaghan, & Madaus, 1977; Fleming & Chambers, 1983; Shulman, 1980; & Haertel, 1984); (Cited in McMillan & Workman, 1998). Essays were less than 1 % of test items and few test questions required students to apply their learning. These studies were done before mandatory standardized testing, the movement towards authentic assessment, and most contemporary school reform.

Stiggins and Conklin, though stressing high school classroom assessment, reported one of the first large-scale studies in 1992. They complained about the dismal state of assessment by citing previous research including Brookhart (1997a) who claimed that teachers spend one-third to one-half of their professional time engaged in either formal or informal assessment activities; Door- Bremme & Herman (1982) who concluded that most assessments were teacher developed and teacher observations were a frequent form of assessment; Gulickson (1982) who provided evidence of lack of quality control strategies; and Shulman (1980) who found that most teachers did not use the results of assessment for any purpose other than assigning grades.

To add depth to the existing knowledge about classroom assessment, Stiggins and Conklin (1992) run a widespread survey among a stratified sample of volunteer teachers throughout the U.S. and found that: 47 % of teachers used teacher-made objective tests, 39 % use published tests and, 57 % use performance assessment. Teachers used these measures for diagnosing, grouping, grading, evaluating, and reporting in various degrees. They relied primarily on teacher made tests (32-48 %), followed by performance assessments (29-34 %) and used published tests the least (9-13 %). Three-fourths (3/4) of teachers had concerns about their own tests in terms of effectiveness, quality, and relevance.

As research progressed through the 1990's more knowledge about teachers' beliefs and practices was gleaned. Researchers attempted to measure more assessment related variables like: criteria that teachers used for selecting assessment methods, the quality of their assessments, feedback on assessments to students, incorporation of higher order thinking, and utilization for grading purposes. Frary, Cross, & Weber (1993) reported that teachers used a variety of assessment approaches with selected response as the most common (71 %), followed by performances (38 %), and essays (37 %). Most teachers used the assessment results to rank students rather than demonstrate mastery of a subject.

Plake and Impara (1993, 1997) measured teachers' knowledge of assessment through a national survey of 555 teachers. The areas of highest knowledge and ability were choosing and scoring assessments. Lowest ratings were in using and communicating assessment results. Overall performance on the survey was 66 % demonstrating empirical evidence of the mournfully low levels of assessment competency. These beliefs about teachers' lack of knowledge and consistency in classroom assessment were supported by Cizek, Fitzgerald and Rachor (1996) who found that teachers' assessment practices were highly variable and unpredictable from characteristics such as gender, years of experience, or grade level among the 143 teachers surveyed. Some teachers (54 %) reported giving a major test or assignments about once every two weeks with the rest giving them less frequently. 75 % said they gave minor assignments at least once per week. Others gave them less frequently. Most (74 %) developed their own assessments. On the average, respondents used 24 grades when calculating final grades. Other factors that teachers took into consideration when determining grades were difficulty of the test (35 %), how the class

performed (43 %), the individual student's ability (51 %), and the individual student's effort (42 %). However, it was puzzling to observe that teachers stated that they lacked training in assessment but then developed their own assessments.

Contrary to other research, Bol, Stephenson and Nunnery (1998), in a study including 893 teachers, were concerned with measuring the influences of teaching experience, grade level, and subject area on classroom assessment practices. They asked the participants about their frequency of use of various assessments, their preparation in developing and administering them, and their beliefs about how accurately various methods reflected student achievement. The results showed that teachers used observations and alternative assessments more than traditional methods and felt that these measures were more valid measures of student achievement. They also found that more experienced teachers used alternative methods of assessment more often than less experienced teachers.

In a similar study, Mertler (1999) used a stratified (male-female, elementary-secondary, years of experience) sample of 625 teachers. Mertler found that female teachers used alternative techniques more than males. No differences were found based on years of experience. Teachers rarely used statistical analysis of classroom assessment and only 25 % took steps to insure validity and reliability. Although the study purported to be examining assessment practices it was more specifically aimed at differences in assessment practices between sub-groups.

Several research studies have been conducted on classroom grading practices. They have provided a number of illuminations concerning teachers' use of assessment rubrics. Stiggins, Frisbee and Griswold (1989) analyzed grading practices in relation to the AFT, NEA, and NCME standards. Through interviews and observations they found that high school teachers use a wide variety of approaches to grading. They want grades to reflect not only academic achievement and effort, but also to motivate students. Contrary to recommended practice, it was found that teachers value student motivation and effort, and set different levels of expectation based on ability. They recommended further research on alignment between assessment policies and practices.

Rubrics can be effective if designed and used properly, but fraught with problems if not (Andrade 2000). Aschbacher (1999), Wenzlaff, Fager and Coleman (1999) and Andrade (2000) studied elementary and high school teachers' use of rubrics, comparing practices to guidelines. They concluded that rubrics were often not aligned with goals, the criteria were often unclear, the level of intellectual challenge was limited, and teachers had little training in their design and use.

Loyd and Loyd (in Phye, 1997) state that "it is not surprising that there are as many grading practices and systems as there are teachers" (p. 481). They suggest principles of effective grading which include the following: grading expectations for performance and achievement need to be clear and explicit with clear descriptors of the targets; communication of these expectations must be clear to the student as well as other stakeholders such as parents and administrators; all students need to be treated fairly and the expectations need to be reasonable; and grading should support, enhance, and inform the instructional process. Gallagher (1998) adds to this list the importance of aligning instructional strategies with grading policies and the use of valid and sufficient data collected over time.

After reviewing classroom assessment (practices) research as a way of setting a well-documented theoretical foundation for the actual study, there a noticeable need to review the existing documents about the standards and recommended practices of assessment.

6. 3. Research on Alternative Assessment

In the 1990's there was backlash to the perceived limitations of standardized tests¹². Critics of these tests state that they are superficial, do not measure the wide range of standards that schools are expected to meet, do not improve

¹²**Standardized Tests** are "tests with uniform procedures for administration and scoring and often allow a student's performance to be compared with the performance of other students at the same age or grade level on a national basis.... They can serve a number of purposes: provide information about students' progress, diagnose students' strengths and weaknesses, provide evidence for placement of students in specific programs, provide information for planning and improving instruction, help administrators evaluate programs, and contribute to accountability (Santrock 2008: 532-533). They are ordinarily constructed according to one of two models: norm-referenced or criterion-referenced. For the former, a distribution of test scores is established for a reference population. New scores are typically presented in conjunction with the corresponding percentile with respect to that reference distribution. For the latter, two or more ordered categories are defined in terms of fixed thresholds on the score scale, and a new score is labeled in terms of the category into which it falls. Standardization is a prerequisite for fairness when scores must be comparable. It demands, at a minimum, that the tests be administered under uniform conditions and graded according to a fixed set of rules or rubrics. The degree of standardization is particularly important for school-leaving and selection examinations—and varies considerably from country to country and among regions within a country. In the past, standardization has been best achieved for examinations set by testing agencies, such as the University of Cambridge Local Examination Syndicate that operate internationally. External examinations, set by regional or national authorities, are almost always standardized.

instruction, and are often punitive (Linn & Gronlund, 1995; Kohn, 2001; Phe, 1997; Popham, 2002); they include only emphasis on recall of factual knowledge rather than higher-order problem solving and, and misalignment between these methods and understanding of cognition and learning (Shepard, 1997). Shift was made to introduce more alternative assessment measures into language classrooms. Emphasis was placed on creating more authentic assessments (Burstein, 1994; Linn, Baker & Dunbar, 1991; McTighe, 1996; Wiggins, 1990). There was speculation that alternative forms of assessment met other educational reform goals such as applying learning to other contexts, encouraging higher level thinking, and demonstrating problem solving. However, most of the published research simply made recommendations for policy and procedures (Aschbacher, 1993; Kane, Khattri, Reeve, & Adamson, 1997; Shepard, 1997).

There is some confusion and disagreement regarding alternative forms of assessment that have also been referred to as performance assessment or authentic assessment. Bracey (2000) describes authentic assessment as an attempt to measure a performance directly in a 'real-life' setting. Wiggins (1998) adds to this by stating that "assessment is authentic when we anchor testing in the kind of work real people do, rather than merely eliciting easy-to-score responses to simple questions" (p.21). The National Education Association (NEA, 1993 cited in Petty 2001) states that "an authentic assessment engages students in challenges that closely represent what they are likely to face as everyday workers and citizens. The context, purpose, audience, and constraints of an authentic assessment must connect in some way to real situations and problems." According to the National Center for Research on Evaluation, Standards, and Student Testing (NCRESST) "alternative assessment is also called authentic or performance assessment and requires students to generate a response to a question or a problem rather than choose from a set of responses provided to them" (NCRESST: 2001). Linn and Gronlund (1995) attempted to demystify and clarify some of this confusion:

Performance assessments are frequently referred to as 'authentic assessment' to emphasize that they assess performance while students are engaged in problem solving and learning experiences that are valued in their own right, not just as a means of appraising student achievement. However, not all performance assessments are 'authentic' in the case that they engage students in solving real problems. (p. 13)

Performances require students to "actively accomplish complex and significant tasks while bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems" (unknown source, taken from a document from a 2002 workshop). Rudner and Shafer (2002, p. 65) explain performance-based assessment as a "set of strategies for the application of knowledge, skills, and engaging to work habits through performance of tasks that are meaningful and engaging to students". Performances generally have pre-established standards that students are striving towards. Airasian (1994) describes distinguishing characteristics of performances as a student demonstration of learning that can be broken down into smaller steps, is directly observable, and measured by performance on the smaller steps.

Many authors embrace the use of alternative forms of assessment (Brualdi, 1998; Glatthorn, 1998; McTighe, 1996). Wiggins (1990) says that "a move toward more authentic tasks and outcomes improves teaching and learning: students have greater clarity about their obligations and teachers come to believe that assessment results are both meaningful and useful for improving instruction" (p. 2). Eisner (1991) asserts that "performance assessment is a closer measure of our children's ability to achieve the aspirations we hold for them than are conventional forms of testing" (p. 1).

As previously defined, authentic assessment measures are performances that relate to real situations and problems. These performances can include demonstrations, exhibitions, products, and portfolios. The key factors in their use as assessment are that they measure learning in valid and reliable ways. They align with expected outcomes, and have measurable components. Research on authentic assessment has explored various aspects including design, scoring, effects on teaching and learning, professional development, validity, reliability, and costs. Those are relative to authentic assessment (used interchangeably with performance assessment) in the classroom and will be reviewed.

CRESST has been one of the primary sources for research on performance assessment. Baker, Aschbacher, Nieme, and Sato (1996) summed up five years of research at the secondary level by Linn, Baker, Burstein, Dietel, Shepard, Aschbacher, Herman and others. The research led CRESST to develop valid scoring techniques, reduce score variability, and strengthen validity criteria.

In 1997 the U.S. Department of Education's Office of Educational Research studied 16 schools to clarify the nature and effects of performance assessment (Kane, Khattri, Reeve, & Adamson, 1997) and found that schools had multiple purposes in using performance assessment, that forms and formats varied tremendously, and that teachers used different types of scoring methods stressing that the efficacy of using performance assessments lies in its potential to raise both students and teachers motivation to learn through research projects and other performance-based assignments than they are with other types of assignments (p. 4). They conclude by saying there is sometimes lack of clarity in performance measures and standards. To improve its value they recommend validity and reliability be strengthened, rubrics be used, and adequate resources, professional development, and support be provided.

Below is a presentation of additional studies that have focused specifically on classroom use of authentic assessment, though they remain limited in scope.

Borko, Flory, and Cumbo (1993) found that while teachers incorporated authentic assessments in the classroom they did not make any other fundamental changes in their curriculum or instruction. Aschbacher (1993), studying the barriers and facilitators for teachers rather than student outcomes, found the biggest problems and barriers were a focus on learning activities rather than student outcomes, difficulties specifying criteria for judging student work, and lack of time to plan, practice, use, and reflect. Shepard (1997) affirmed that teachers need extensive time, training, and assistance to be able to use alternative forms of assessment effectively.

Suurtamm (2000) found that support throughout the educational community needs to be in place to facilitate a change from traditional to authentic assessment in the classroom. Similarly, Soodak and Martin-Kniep (n.d) found that teachers need to see a strong link between curriculum, instruction, and assessment in order to facilitate change.

Performance assessment is frequently measured through the use of rubrics. These rubrics are "scoring guides with specific pre-established performance criteria" (Mertler, 2001). "They clearly define what a range of acceptable and unacceptable performance looks like" (Wenzlaff, Fager, & Coleman, 1999, p.1). "Its purposes are to give students informative feedback about their work in progress and to give detailed evaluation of their final product"

McTighe (1997) feels that performance assessments "are better suited than traditional measures to measure what really counts: the application of knowledge, skills, and understanding in important real-world contexts" (p.1). He offers six principles for performance-based assessment.

- Establish clear performance targets that are linked to instructional goals.
- Strive for authenticity in products and performances.
- Publicize criteria and performance standards.
- Teach, model, and guide students through the strategies to be used.
- Use on-going assessments for feedback and adjustment.
- Document and celebrate progress.

Amy Brualdi (1998) states that "the benefits of performance-based assessments are well documented, ... but teachers are hesitant to use them because they don't often know enough about them" (p.2). Using the work of Airasian, Popham, Stiggins, and Wiggins she offers guidelines for teachers.

- Define the purposes of the performance-based assessment (what type of knowledge or skill and at what level.)
- The activity, whether formal or informal, must take into account time, resources, and amount of data needed.
- The criteria for evaluation must be clearly defined, identify the important components, and be observable and measurable.
- Rate the performance based on a rubric that reflects levels of achievement of each criterion.

Despite the large amount of support for alternative forms of assessment, Haertel (1999) cautions that, "regardless of the value of performance assessments in the classroom, a measurement-driven reform strategy that relies on performance assessments to drive curriculum and instruction seems bound to fail" (p.62). Any measurement driven accountability needs to have appropriate standards, adequate teacher preparation, limited extraneous demands and requirements, and sufficient resources. He argues that since large-scale performance assessment has not "effected sweeping education reform, it should not detract from their value for instructional purposes" (p. 63).

6. 4. Research on Formative Assessment

Both large-scale and classroom assessment can contribute to informing and formatively improving instruction. Formative assessment refers to the ongoing assessment whereby teachers actually gather evidence- feedback- used over the course of instruction to adapt the teaching in ways that would meet students' needs and diagnose their progress toward a long-term objective (Black & William, 1998; Boston, 2002). It is usually collected by different means such as assignments, pretests, midterm student conferences (or tests), theses and projects, oral or comprehensive exams. It as well intends to provide information on "what, how much, and how well students are learning" (Angelo & Cross, 1993, p. 5). A language teacher uses assessment results formatively to (1) review student work by identifying their weaknesses and strengths with regard the objectives set for instruction, and not only to give final marks or grades, (2) increase communication and collaboration between instructor and students by tracking and improving student learning on the basis of collecting feedback on effectiveness of and student satisfaction with classroom activities and teaching, (3) promote the course of teaching and learning to meet students' needs by evaluating curriculum goals, instructional strategies, program and instructional effectiveness, and finally (4) ensure some necessary changes and potential interventions that need be made, during the instructional process,

before the end (summative) results are gathered (Harrington & Reid, 1996; Murphy & Harrold, 1997; Steadman, 1998; Black & William, 1998; Dyck, Pemberton, Woods, & Sundlbye, 1998; Cresap & Eklund, 2000; Hatfield & Gorman, 2000; Fewter & McMillan, 2002). Stiggins (2002) explains that formative assessment also serves to motivate students more effectively than summative. Such feedback helps students identify gaps between their current knowledge and skills and the desired goal as well as to identify specific areas in need of improvement. It stands in contrast to summative assessment, which generally takes place after a period of instructions and requires making a judgment about the learning that has occurred.

Black and William (1998) conducted a review of 250 international journal articles, books, and research to determine whether formative assessment raises academic standards in the classroom. They concluded that "firm evidence shows that formative assessment is an essential component of classroom work and that its development can raise standards of achievement" (p. 139). This was particularly true for low achieving students and those with special needs. Using multiple research studies they summarized primary research findings that are similar to conclusions drawn by other research on classroom assessment. They found an emphasis on rote and superficial learning, a focus on quantity rather than quality, overemphasis on grading rather than learning outcomes, and approaches that result in comparison and competition.

Based on this research, they offer suggestions for improving formative assessment at all grade levels. "Feedback to any pupil should be about the particular qualities of his or her work, with advice on what he or she can do to improve, and should avoid comparisons with other pupils" (Black & William, 1998, p. 145). They encourage self-assessment that provides clear learning targets. This must be combined with an understanding of the goal, evidence about the present position, artifacts of achievement, and strategies for closing the gap. One of their recommended strategies is the use of meaningful, focused dialogue and thoughtful, reflective questioning.

Stiggins (1999b) distinguishes between summative assessment, calling it assessment *for* learning, and formative assessment, calling it assessment *of* learning. If assessments of learning provide evidence of achievement for public reporting, then assessments for learning serve to help students learn more. He describes the basic principles of assessment *for* learning as follows. Teachers understand and articulate, in advance of teaching, the achievement targets that their students are to hit. They inform their students about those learning goals in terms that students understand from the very beginning of the teaching and learning process. Teachers are assessment literate and thus are able to transform those expectations into assessment exercises and scoring procedures that accurately reflect student achievement. They use classroom assessment to build student confidence in themselves as learners, helping them take responsibility for their own learning so as to lay a foundation for life-long learning. Classroom assessment results are consistently translated into informative (not merely judgmental) feedback for students, providing them with specific insights as to how to improve. Students work closely with their teacher to review assessment results, so as to remain in touch with, and thus feel in charge of, their own improvement over time. Teachers continuously adjust instruction based on the results of classroom assessments. Students are actively involved in communicating with their teacher and their families about their achievement status and improvement.

Contrary to the views advocating the beneficial use of formative assessment, an emergent movement has been towards a balance of standardized and classroom assessment that can serve to provide a more impartial approach to assessment, afford it more credibility, and facilitate its use in more constructive ways (Coladarci, 2002; Rabinowitz, 2001). Baker, Linn, Herman, & Koretz (2002); Claycomb & Kysilko (2001); Hanushek & Raymond (2002), as well as national organizations, have made recommendations for the design of these systems, but as Hanushek notes: "it still represents a young and highly selective body of work (p. 1). What is important about this movement is that there is a growing belief "if we wish to take full advantage of the power of assessment to maximize student achievement ... We must rely on a balanced combination of high quality standardized assessments of learning and high quality classroom assessment for learning (Stiggins, 2002)". As the forces on assessment have evolved from standardized, to alternative (authentic), this emerging movement towards balanced assessment systems has the potential to have a significant impact on policies regarding school-based assessment (Cutlip, 2003; Rabinowitz, 2001).

Stiggins (2002a) believes that it is flawed to think that all assessment informs decisions and motivates learning. Standardized assessments of learning provide evidence of achievement for public reporting. He advocates for the use of assessment for learning which he says takes formative assessment to the next level by involving students in the assessment process. Assessment for learning keeps students learning and grants them more confidence and ability to continue to learn at productive levels if they keep trying to learn. He feels that many of his principles of assessment for learning are embedded in the standards for assessment of many national organizations but that teachers haven't necessarily "mastered those essential classroom assessment competencies" (p. 763). Stiggins (2002c, p. 2) states that "there are no good arguments against balancing our assessment of and for learning" and that "harm arises directly from our failure to balance our use of standardized tests and classroom assessment in the service of school improvement" (2002a). Balance benefits all stakeholders including students, teachers, parents, administrators, and the community. A number of well-established organizations were reviewed offering a set of useful quality control criteria for effective classroom assessment. This is what will be the concern

of the subsequent section.

7. Quality Control Criteria for Effective Classroom Assessment

It is important to examine what is meant by effective classroom assessment if one is to get the appropriate tools to evaluate the teachers' actual practices of classroom assessment. Whatever mode of assessment is developed, chosen, adopted, adapted for one's classroom assessment practice, it is mandatory that these should (1) follow a theoretically founded model of how tests are developed, validated and evaluated and (2) aligned with the quality control criteria recommended by the standards of the best practices of classroom assessment so that they can effectively fulfill the functions for which they are used.

The assessment literature contains a generous number of recommendations as to the desired characteristics of assessment. It is the consensus that effective and high quality EA includes both formative and summative practices (Burke, 1999). A number of assessment references were scrutinized in an attempt to build an easily adoptable model for the language testing process as well as a self-contained and comprehensive framework for what might serve as (best) model for a high quality assessment. Sifting through seminal reference books published by famous language testing scholars were reviewed and prompted my adoption of Luoma's (2001) test development and validation process as it has proven to be representing almost all the components of the language test development, validation and evaluation process. In the same vein, nine international organizations¹³ recognized for their potent contributions to setting standards and recommendations for best practices of classroom assessment were resorted to in order to construct a thoughtful structure of what constitutes an effective classroom assessment. Common to all of them are the importance of clear standards and targets for all assessments, multiple measures, monitoring of student progress, and professional assessment literacy. In addition, communication of results, use of assessment data to guide and inform instruction, and management of data are also important. The following is a summary of the model of the language test development and validation adopted and the recommendations for components of effective assessment that need be followed as quality criteria that every language teacher should cater for while developing or choosing an assessment to use to assess his student learning.

For an effective monitoring system that provides charted data, Fuchs and Fuchs (1996) described some guiding criteria for high quality assessment. A few of these are:

- Measurement of important learning outcomes;
- Measurement that addresses important educational decisions;
- Measurement that is compatible with a variety of instructional models;
- Clear descriptions of student performance that can be linked to instruction;
- Easy administration, scoring, and interpretation;
- A process that communicates learning goals to teachers and students; and
- A system that generates accurate and meaningful information.

Other scholars believe that classroom assessment should reflect "real-life" (i.e., outside of the classroom) tasks and require students to utilize higher order thinking skills (Crotty, 1994; Leon & Elias, 1998) to fulfill on-demand duties and tasks. This gave birth to another assessment approach known as contextual (authentic) assessment (CAA) advocating that these assessments should:

- require intellectually worthy tasks,
- mirror best instructional activities, and
- consist of ill-structured challenges that are similar to the complex ambiguities of life (Wiggins, 1990),
- assess a wider range of learning outcomes,
- use a wider range of types of assessment tasks, and

¹³ I. NEASC: New England Association of Schools and Colleges,

II. Joint Committee on Standards for Educational Evaluation that was founded in 1975 to develop standards for educational evaluation. It originally initiated by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, the American Association of School Administrators, Association for Supervision and Curriculum Development, Consortium for Research on Educational Accountability and Teacher Evaluation, The National Association of Secondary School Principals, the National School Boards Association, and the Council of chief State School Officers.

III. The American Federation of Teachers, National Council on Measurement in Education, National Education Association.

IV. The Committee on Instructionally Supportive Assessment including 5 organizations: The American Association of School Administrators, The National Association of Elementary School Principals, The National Association of Secondary School Principals, the National Education Association, and the National Middle School Association.

V. National Forum on Assessment.

VI. National Education Association.

VII. Local Assessment System by Theodore Coladarci.

VIII. WestEd: Assessment and Standards Development Services written by Stanley Rabinowitz.

IX. Assessment Training Institute founded by Richard Stiggins in 1992.

- assess in more authentic contexts (Bell & Cowie, 2001)

Luoma (2001) contributed to this issue with a well-stratified model for the process of test development and validation. Below is a summary of its components and the complete skeleton of this process. According to Luoma (2001) a well-designed test should:

- Use well-developed test specifications operationalizing the construct assessed.
- Use well-developed tasks and assessment criteria.
- Pilot the test.
- Revise the test material (tasks, scoring rubrics and criteria).
- Use well-structured administration procedures.
- Validate the test by evaluating test product against theory (construct) and linking it to future validation research.
- Operationalize the test by developing new tasks.
- Operationalize test administration.
- Monitor test reliability and maintain it for future use.
- Revalidate the test as a part of the continuous validation process.

8. Conclusion

The aforementioned review attempts to give a general overview of classroom assessment practices / research in response to one of the higher demands of practicing teachers as to how to understand the working mechanisms of such a detrimental task to which they have to respond with utmost degrees of accountability. Further practical details and sample test designs will be provided in the forthcoming publications taking novice and in-service teachers through the intricate stages of the process.

Acknowledgements

I thank Dr. Redouane Madaoui, a faculty colleague, for his meticulous editing of an old version of the current article. Thanks are too due to my MA students in the master program of “Applied Linguistics & Language Studies”, who are graduating this year, for having responding to the high demands and requirements of the training in the Language Testing Course I was teaching them during the 2015-2016 winter semester. Their strong affiliation and high degree of responsiveness to the tasks and reflections on the “Educational Assessment (Testing) Reflective Journal” and the ongoing take-off assignment as well as their indulgence in the test design process according to the Bachman & Palmer’s (2010) updates in the the test design process. That has spiced up my appetite to go through the challenges of self-enlightenment and erudition, sharing, educating and training in the area of “Educationol Assessment”. Heartfelt thanks are expressed to the anonymous reviewers who have validated the quality of the current article.

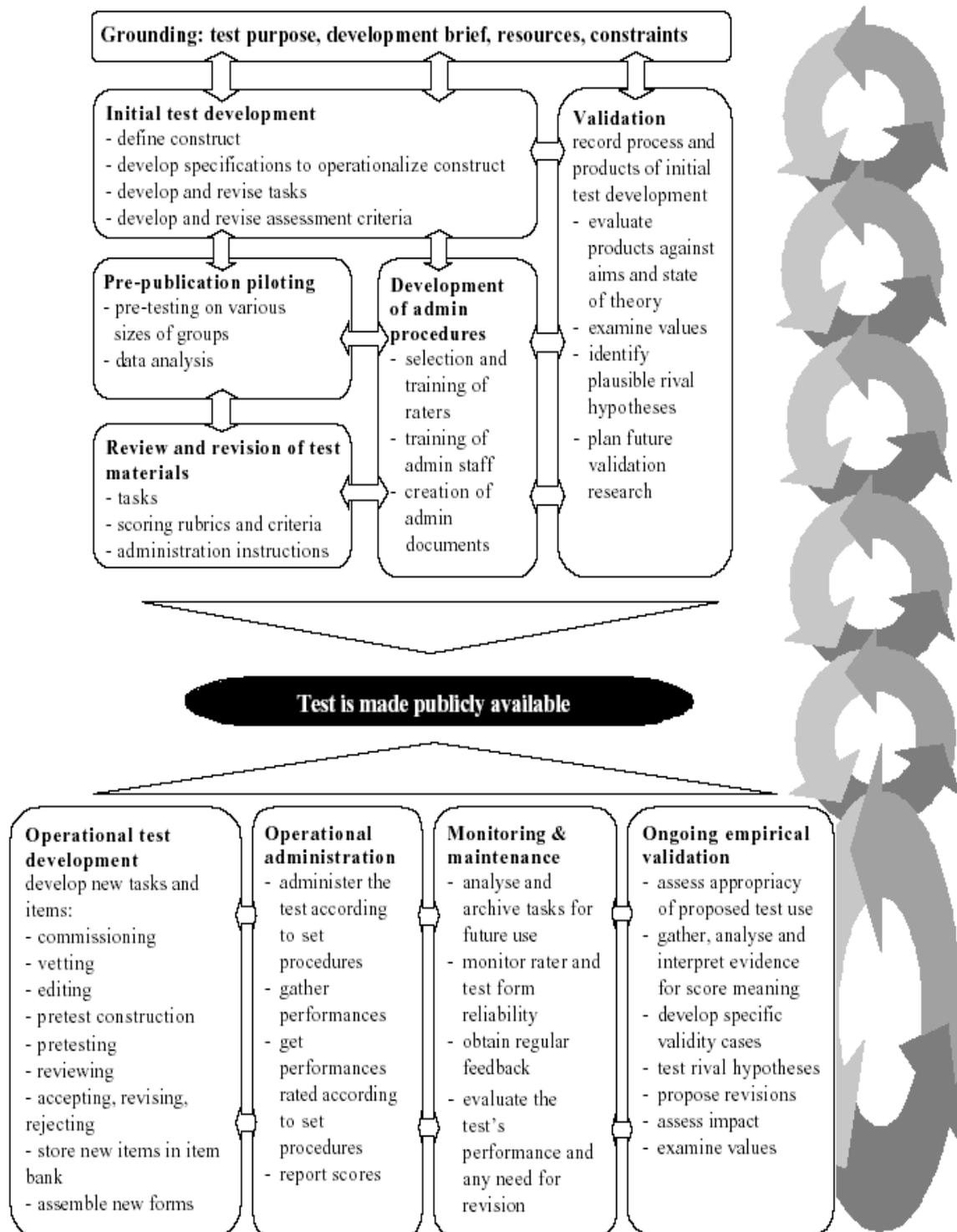


Figure 2. The test development and validation process. Adopted from Luoma 2001: 150.

References

- Airasian, P. W., Kellaghan, T., & Madaus, G. F., (1977). The stability of teachers' perceptions of pupil characteristics. *Irish Journal of Education*. 11(2), 74-84.
- Airasian, P. W. (1997). *Classroom assessment*. (3rd ed.). New York: McGraw-Hill.
- Airasian, P. W. (1984). Classroom assessment and educational improvement. Paper presented at the conference Classroom Assessment: A key to educational Excellence, Northwest Regional Educational Laboratory, Portland; Oregon.
- Airasian, P.W. (1994). *Classroom assessment*. New York: McGraw-Hill.

- Alderson, J.C. (1990) Testing Reading Comprehension Skills (Part One), *Reading in a Foreign Language*, 6(2), 425-38.
- Alderson, J.C. (1990) Testing Reading Comprehension Skills (Part Two) Getting Students to Talk about Taking a Reading Test (a Pilot Study), *Reading in a Foreign Language*, 7(1), 465-503.
- Alderson, J. C. (2000 a). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C., & Banerjee, J. (2001). State of the art review : language testing and assessment (part two). *Language Teaching*, 35(2), 79-113.
- Alderson, J. C. & North, B. (Eds.). (1991). *Language testing in the 1990s*. London: Macmillan.
- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 1 3-1 8.
- Angelo, T. A., & Cross, P. K (1993). *Classroom assessment techniques: a handbook for college teachers* (2nd ed.). San Francisco: Jossey-Bass.
- Aschbacher, P. R. (1991a, June). What have we learned from writing assessment that can be applied to performance assessment in other areas? Presentation to the ECS/CDE Assessment Conference, Breckenridge, CO.
- Aschbacher, P. R. (1991b). Performance assessment: State activity, interest, and concerns. *Applied Measurement in Education*, 4(4), 275-288.
- Aschbacher, P. R. (1 993). *Issues in innovative assessment for classroom practice: Barriers and facilitators*. (CSE Technical Report 359). National Center for Research on Evaluation, Standards, and Student Testing. University of California, Los Angeles.
- Aschbacher, P. R. (1 999). *Developing indicators of classroom practice to monitor and support school reform*. (CSE Technical Report 5 1 3). National Center for Research on Evaluation, Standards, and Student Testing. University of California, Los, Angeles.
- Astin, A. W. (1991). *Assessment for Excellence: the philosophy and practice of assessment and evaluation in higher education*. New York: Macmillan.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. (1990) Assessment and Evaluation, 210-226. *Annual Review of Applied Linguistics*, 10:
- Bachman, L.F. (1991) What Does Language Testing Have to Offer? *TESOL Quarterly*, 25(4), 671 -704.
- Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly*, 19, 535-56.
- Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Baker, R. (1996). Language testing and the assessment of dementia in second language settings: A case study. *Language Testing*, 13, 3-22.
- Baker, E. L., Aschbacher, P. R., Niemi, D., & Sato, E. (1992). CRESST Performance assessment models: Assessing content area explanations. Los Angeles: CRESST.
- Baker, E. L., Ashbacher, P. R., Neimi, D., & Sato, E. (1996). *Performance assessment models: Assessing content area explanations*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.
- Baker, E. L., Linn, R. L., Herman, J. L., & Koretz, D. (2002). *Standards for educational accountability systems* (Policy Brief No. 5). Los Angeles, CA: University of California, Center for the Study of Evaluation.
- Banta, T. W. et al. (1993). *Making a difference: Outcomes of a decade of assessment in higher education*. San Francisco: Jossey-Bass.
- Banta, T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (Eds.). (1996). *Assessment in practice: Putting principles to work on college campuses*. San Francisco: Jossey-Bass.
- Banergee, J. & Luoma, S. (1997). Qualitative approaches to test validation. In C.Clapham & D. Corson (eds), *Encyclopedia of language education*, Vol. 7: Language testing and assessment, pp: 275-287. Amsterdam: Kluwer.
- Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht: Kluwer Academic Publishers.
- Black, P., & William, D. (1998). Assessment and classroom learning. *EA: Principles, Policy and Practice*, 5(1), 7-74
- Black, P., & William, D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan* 80(3), 139-148
- Bol, L., Stephenson, P.L. & Nunnery, J. A. (1998). Influence of experience, grade level, and subject area on teachers' assessment practices. *The Journal of Educational Research*, 91(6), 323-330.
- Borko, H. Flory, M & Cumbo, K (1993). Teachers' ideas and practices about assessment and instruction: A case study of the effects of alternative assessment in instruction, student learning, and accountability practices. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA,

- April 12-16, 1993.
- Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research, & Evaluation*, 8(9). Retrieved January 19, 2011 from <http://PAREonline.net/getvn.asp?v=8&n=9>.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12. doi:10.1111/j.1745-3992.2003.tb00139.x
- Brookhart, S. M. (1997a). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education*, 10, 161-180
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall regents.
- Brown, T.L.G. (2004). Teachers' conceptions of assessment: implications for policy and professional development. *Assessment in Education*, 11 (3), 305-322.
- Brualdi, A. (1998). Implementing performance assessment in the classroom. *Practical Assessment, Research & Evaluation*, 6(2). Retrieved on March 3, 2008, from <http://pareonline.net/getvn.asp?v=6&n=2>
- Burstein, L. (1994). *Performance-based assessment for accountability purposes: Taking the plunge and assessing the consequences*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Chun, C. (2002). Looking where the light is better: A review of the literature on assessing higher education quality. *Peer Review*, Winter/Spring, 16-25. Retrieved June 13, 2007, from <http://www.aacu.org/peerreview/pr-sp02/pr-sp02feature3.cfm>
- Cizek, G. J. (1991, April). *The effect of altering the position of options in a multiple-choice examination*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Cizek, G. J. (1991, May). Innovation or inervation? Performance assessment in perspective. *Phi Delta Kappan*, 72 (9), 695-699.
- Cizek, G. J. (2001a). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(4), 19–27.
- Cizek, G. J. (2003a). *Detecting and integrating classroom cheating: Promoting integrity in schools*. Thousand Oaks, CA: Corwin.
- Cizek, G. J. (2003b). Educational testing integrity. In J. E. Wall, & G.R. Walz (Eds.) *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 363-386).
- Cizek, G. J., Rachor, R. E., & Fitzgerald, S.M. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159-179
- Clare, L. (2000). *Using teacher's assignments as an indicator of classroom practice*. (CSE Technical Report 532). National Center for Research on Evaluation, Standards, and Student Testing. University of California, Los Angeles.
- Claycomb, C., & Kysilko, D. (2001). *The purposes and elements of effective assessment systems*. National Association of State Boards of Education. Study Group on Statewide Assessment Systems.
- Coldarci, T. (2000). *Measured Measures: Technical considerations for developing a local assessment system*. Maine Comprehensive Assessment System, Technical Advisory Committee. Maine Department of Education.
- Cowart, S. C. (1990). *A survey on using student outcomes measures to assess institutional effectiveness*. Iowa City, IA: American College Testing Program.
- Cromey, A., & Hanson, M. (2000). An exploratory analysis of school-based student assessment systems. North Central Regional Educational Laboratory. Retrieved April 6th, 2003 from <http://www.ncrel.org/re/eval/mi/analysis.htm>
- Crotty, E.(1994). The Role of Cooperative Learning in Authentic Performance Assessment Approach. *Social Science Record*, 31(1), 38-41.
- Cutlip, G. W. (2003). *Balanced assessment: The key to accountability and improved student learning*. Washington, D. C.: National Education Association.
- Dorr- Bremme, D., W. & Herman, J. (1982). *Testing and Assessment in American Public Schools: Current Practices and Directions for Improvement*. Research into Practice [Project], Test Use Monograph.
- Dyck, N., Pemberton, J. Woods, Sundbye, N. (1998). *Creating inclusive schools: a new design for ALL students*. Lawrence, KS: curriculum solutions.
- Eisner, E. (1999). The uses and limits of performance assessment. *Phi Delta Kappan*. Retrieved on December 18, 2002 from <http://www.pdkintl.org/kappan.htm>
- El-Khawas, E. (1988). *Campus trends 1988*. Higher Education Panel Report No. 77. Washington, DC: American Council on Education.
- El-Khawas, E. (1989b). How are assessment results being used? *Assessment Update*, 1(4), 1-2.
- El-Khawas, E. (1990). *Campus trends 1990*. Higher Education Panel Report No. 80. Washington, DC: American Council on Education.
- El-Khawas, W. (1995). *Campus trends 1995*. Higher Education Panel Report No. 85. Washington, DC: American

- Council on Education
- Embretson, S. E., & Hershberger, S. L. (1999). *The new rules of measurement: What every psychologist and educator should know*. Mahwah, N.J.: L. Erlbaum Associates.
- Ewell, P. T. (1988a). Implementing assessment: Some organizational issues. In T. W. Banta (Ed.), *Implementing outcomes assessment: Promise and perils* (New Directions for Institutional Research No. 59, pp. 15-28). San Francisco: Jossey-Bass.
- Ewell, P. T. (1988b). Outcomes, assessment, and academic improvement: In search of usable knowledge. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. IV, pp. 53-108). New York: Agathon Press.
- Ewell, P. T. (1997). Strengthening assessment for academic quality improvement. In M. W. Peterson, D. D. Dill, L. A. Mets, and Associates (Eds.), *Planning and management for a changing environment: A handbook on redesigning postsecondary institutions* (pp. 360-381). San Francisco: Jossey-Bass.
- Finocchiaro, M. and Sako, S. (1983). *Foreign language testing- a practical approach*. New York: Regents Publishing Co.
- Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice*, 12(3), 23-30.
- Fuchs, L.S, & Fuchs, D. (1996). Combining performance assesment and curriculum-based measurement to strengthen instructional planning. *Learning disabilities research & Practice*, 11(3), 183-192.
- Gallagher, J. D. (1998). Classroom assessment for teachers. Upper Saddle River, New Jersey: Prentice-Hall.
- Gulickson, I. R. (1982). Survey data collected in survey o South Dakota teachers' attitudes and opinions toward testing. Vermillion: University of South Dakota.
- Gray, P. J., & Banta, T. W. (Eds.). (1997). *The campus -level impact of assessment: Progress problems and possibilities*. (New Directions for Higher Education No. 100). San Francisco: Jossey-Bass.
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching*. (6th ed.). New York: Macmillan.
- Linn, R. L. & Gronlund, N. E. (1995). *Measurement and assessment in teaching, Seventh edition*. New York: Macmillan.
- Haertel, E. H. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement*, 8, 333-346.
- Hatfield, S. R., & Gorman, K. L. (2000). Assessment in education-the past, present, and future. Assessment in Business Education National Business Education Association *Yearbook*, 38, 1-10. Reston,VA: National Business Education Association.
- Huba, M.E. & Freed, J.E. (2000). Learner-centered assessment on college campuses: Shifting the focus from teaching to learning. Needham Heights, MA: Allyn & Bacon.
- Johnson, R., Prus, J., Andersen, C. J., & El-Khawas, E. (1991). *Assessing assessment: An in-depth status report on the higher education assessment movement in 1990*. Higher Education Panel Report No. 79. Washington, DC: American Council on Education.
- Kane, M. B., Khattri, N., Reeve, A., & Adamson, R. J. (1997). *Studies of educational reform: Assessment of student performance*. (Contract No. RP 91-172004). Washington DC: U.S. Department of Education. Retrieved September 23, 2001, from <http://www.ed.gov/pubs/SER/ASP/stude.html>
- Kohn, A. (2001). Fighting the tests: A practical guide to rescuing our schools. *Phi Delta Kappan*, 82(5), 349-357.
- Lado, Robert (1961). *Language Testing*. New York: McGraw Hill.
- Leon, S., & Elias, M. (1998). A comparison of portfolio, performance and traditional assessment in the middle school. *Research in Middle Level Education Quarterly*, 21(2), 21-37.
- Linn, R. L. (2001). Assessments and accountability. *Practical assessment, Research, and Evaluation*, 7(11). Retrieved October 16, 2002 from <http://ericae.net>
- Linn, R., Baker, E. L., & Dunbar, S. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
- Linn, R. L. & Gronlund, N. E. (1995). *Measurement and assessment in teaching* (7th ed.). Englewood Cliffs, New Jersey: Prentice Hall.
- Linn, R. L., & Baker, E. L. (2001). *Validity issues for accountability systems (CSE Technical Report 585)*. National Center for Research on Evaluation, Standards, and Student Testing. University of California, Los Angeles.
- Luoma, S. (2001). *WHAT DOES YOUR TEST MEASURE?* Construct definition in language test development and validation. Unpublished PhD. Thesis, University of Jyväskylä.
- Maki, P. (2002). Moving from paper work to pedagogy: Channelling intellectual curiosity into a commitment to assessment. *AAHE Bulletin*, May.
- Marzano, R. J, Pickering, D., & McTighe, J. (1993). *Assessing student outcomes: performance assessment using the dimensions of learning model*. Alexandria, VA.: Association for Supervision and Curriculum Development.

- McDonnell, L. (1994). Assessment policy as persuasion and regulation. *American Journal of Education*, 102, 395-420.
- McMillan, J. H. (2001a). *Classroom assessment: Principles and practice for effective instruction*. (2nd. ed.). Boston, MA: Allyn & Bacon.
- McMillan, J. H. (2001b). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32.
- McMillan, J.H. & Workman, D. J. (1998). *Classroom assessment and grading practices: A review of the literature*. Metropolitan Educational Research Consortium. Richmond, VA. (ERIC Document Reproduction Service, ED453263).
- McTighe, J. (1996). What happens between assessments? *Educational Leadership*, 54(4), 6-12.
- McTighe, J., & Wiggins, G. (1999). *The Understanding by Design handbook*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mehrens, W. A., Popham, J. W. and Ryan, J.M. (1998). How to Prepare Students for Performance Assessments, *Educational Measurement: Issues and Practice*, 3, (3), 18-22.
- Mehrens, W. A. (1992). Using performance assessment for accountability purposes. *Educational Measurement : Issues and Practices*, 11(1), 3-9.
- Mertler, C. A. (1999). Assessment of student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, 120(2), 285- 297.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment; Research, and Evaluation*, 7(25). Retrieved August 2, 2006.
- Mundrake, G. A. (2000). The evolution of assessment, testing, and evaluation. In: Rucker, J. *Assessment in Business Education*, 38, *NBEA Yearbook*. Reston: NBEA.
- Murphy, E. (2000). Strangers in a strange land: Teachers' beliefs about teaching French as a second language in online environments. Unpublished Doctoral dissertation, Universite Laval, Quebec.
- National Council for Accreditation of Teacher Education (2002). *Professional standards Accreditation of Schools, Colleges, and Departments of Education*.
- Nitko, A.J. (1995). Is the Curriculum a Reasonable Basis for Assessment Reform? *Educational Measurement: Issues and Practice*, 14(3), 5-17.
- Nitko, A.J. (1996). *Educational Assessment of Students (2nd ed.)* Engwod Cliffs, NJ: Prentice-Hall/ Merrill Education.
- Nitko, A.J. (2001). *Education assessment of students (3rd ed.)*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Omaggio, A. (1986). *Teaching Language in Context: Proficiency oriented instruction*. Boston, MA: Heinle and Heinle Publishers.
- Ory, J. C., and Parker, S. A. (1989). Assessment activities at large, research universities. *Research in Higher Education* 30, 375-385.
- Pelligrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *knowing what students know: the science and design of EA*. Washington, D.C.: National Academy Press.
- Peterson, M. W., & Einarson, M. K. (2001). What are colleges doing about student assessment: Does it make a difference? *The Journal of Higher Education*, 72(6), 629-669.
- Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment*. (pp. 55-68). New York: Academic Press.
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12, 39.
- Popham, W.J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56(6), 8-15.
- Popham, W.J. (2000). *Assessments that illuminate instructional decisions*. Paper presented at the Conference on Large-Scale Assessment, council of Chief State school Officers, Snowbird, UT.
- Popham, W.J. (2000). *Modern educational measurement: issues and Practice*, 12(4), 10-12, 39.
- Popham, W.J. (2002). *Combating the fraudulence of standards-based assessment*. Presentation at the Large-Scale Assessment Conference of the Council of Chief State Officers.
- Popham, W.J. (2003b). *Test Better, Teach Better: The Instructional Role of Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Rabinowitz , S. M. (2001). Balancing state and local assessments. *School Administrator*, December 2001. Retrieved June 6, 2003 from <http://www.aasa.org>
- Rabinowitz, S. & Ananda, S. (2001). *Balancing local assessment with state-wide testing*. Wested. Retrieved July 6th, 2003 from <http://wested.org>
- Ratcliff,R., van Zandt, T ., & McKoon, G . (1995). Process dissociation, single process theories, and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803-814.
- Rossman, J. E., & El-Khawas, E. (1987). *Thinking about assessment: Perspectives for presidents and chief*

- academic officers*. Washington, DC: American Council on Education and the American Association for Higher Education.
- Rudner, L. M. & Schafer, W. D. (Eds.) (2002). *What teachers need to know about assessment*. Washington, D.C: National Education Association.
- Santhrock, J. w. (2008). *Educational Psychology*. 3rd edition. Mc. Graw Hills Companies. NY.
- Schafer, W.D. & Lissitz, R.W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education*, 38(3), 57-63.
- Schafer, William. (1991). Essential assessment skills in the professional education of teachers. *Educational Measurement: Issues and Practice*, 10, 1-12.
- Shavelson, R. J. & Huang. L. (2003). Responding responsibly to the frenzy to assess learning in higher education. *Change* 35(1): 10-19.
- Shepard, L. A. (1997). *Insights gained from classroom-based assessment project*. (CSE Technical Report 451). National Center for Research on evaluation, Standards, and student testing, University of California, Los Angeles.
- Shepard, L.A, Flexer R., Hiebert, E., Marion, S., Mayfield, V., & Weston, T. (1995). *Effects of introducing classroom performance assessments on student learning* (CSE Technical Report 525). Bolder: University of Colorado, National Center for Research on Evaluation, standards, and student Testing.
- Steadman, M. (1998). Using classroom assessment to change both teaching and learning. *New Directions for Teaching and Learning*, 75, 23-35.
- Stiggins, R. J. (2001a). *The unfulfilled promise of classroom assessment*. Portland, Oregon: Assessment Training Institute.
- Stiggins, R. J. (2001b). *Student-involved classroom assessment*. Upper Saddle River, N.J.: Prentice-Hall Inc.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Stiggins, R. J. (2002b). *Leadership for excellence in assessment-2003: A planning guide to the perfect assessment system*. Portland, OR: Assessment Training Institute
- Stiggins, R. J. (2002c). Assessment for learning. *Education Week*, March 13, 2002. Retrieved on from <http://www.edweek.org>.
- Stiggins, R. J. & Conklin, n. F. (1992). *In teachers' hands: Investigating the practices of classroom*. Albany, NY: State University of New York Press.
- Stiggins. R. (2004). New Assessment Beliefs for a New School Mission. *Phi Delta Kappan*, 86 (1), 22-27
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5-14.
- Walvoord, B., Bardes, B., & Denton, J. (1998). Closing the feedback loop in classroom based assessment. *Assessment UPdate: Progress, Trends and Practices in Higher Education* 10(5), 1–11.
- Wiggins, G. (1998). *Educative assessment*. San Francisco, CA: Jossey-Bass.
- Wiggins, G. (1989a). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Wiggins, G. (1990). The case for authentic assessment. Office of Educational Research and Improvement, U.S. Department of Education. Retrieved from ERIC database.
- Wiggins, G. P., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Wiggins, G. P. (1993). *Assessing student performance*. San Francisco: Jossey-Bass Publishers.
- Wolf, D., Bixby, J., Glenn, J. III., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), *Review of research in education, Vol. 17*. Washington DC: American Educational Research Association.
- Zeliff, N. D., (2000). Alternative assessment. *National Business Education Yearbook 2000*, pp. 91-102.