THE IMPACT OF MIDDLE LEVEL TEACHERS' ASSESSMENT PRACTICES AND

INSTRUCTIONAL LEADERSHIP ON STUDENT ACHIEVEMENT

Review of Related Literature

Kristin Weiser Matthews

Chapter 2

Review of the Related Literature

Introduction

To understand the progression of assessment practices and beliefs, this review of literature was organized into four main sections: (1) historical conceptions about assessment, (2) current conceptions about assessment, (3) strategies for classroom assessment, and (4) instructional leadership. Section one, historical conceptions about assessment, begins with the beginnings of educational assessment and concludes with modern assessment practices in the United States. Section two, current conceptions about assessment, focuses on the research on assessments, classroom assessments and assessment literacy. Section three, strategies for classroom assessment, takes the previous literature on classroom assessment and narrows in on specific strategies listed as best practices for classroom assessment. Section four, instructional leadership, centers on the impact that the principal can have on teachers' assessment practices.

Historical Conceptions about Assessment

Assessment in education has a long tradition in the United States. Actually, there has never been in a time in the history of American public schools without some form of assessment in place (Cuban, 2004). During colonial times, oral examinations were administered to verify that children learned the prescribed material (Congress, 1992). In 1845, Horace Mann convinced the Boston School District to replace oral quizzes with written tests (Gerberich, 1963).

The tests were not intended to be used for comparisons among children and their schools; however, they were quickly utilized for that purpose. Long before the Civil War,

schools used mandated written examinations to assess student progress and to aid in policy

decisions (Congress, 1992). In 1894, J. M. Rice demonstrated that tests of performance could be

important for assessing instruction, curriculum, and student learning. He administered spelling

tests in different school systems, looking to see if added instructional time improved student

performance (Resnick, 1980; Gerberich, 1963).

The Taylorism of the American Education System

The Industrial Revolution brought social efficiency and scientific measurement to

education (Shepard, 2000b; Resnick, 1980). Schools played a critical role in response to the

changing job market brought on by the Industrial Revolution. More white-collar opportunities

and rising standards created more accountability on the part of schools. In addition, schools had

to answer to the acculturation of the largest wave of immigrants in the nation's history. With the

growing population and the changing job market, public schools saw a huge increase in the 14 to

17 year age group in schools (Resnick). Researchers were influenced by public concerns about

education that were shaped by industrialization, fears of loss of community, and the need to

absorb and "Americanize" large groups of immigrants (Shepard, 2000a).

The organizational theorist, Frederick Taylor, developed his theories of scientific

management during this time period. He believed that in order to achieve optimum efficiency,

tasks were split into many separate elements that were to be observed and evaluated. These

models of efficiency greatly influenced the industries on the United States (Morgan, 2006).

Tayloristic influences found their way into the American education system especially

through the work of E. L. Thorndike. Thorndike was an experimental psychologist teaching at

Columbia University (Teachers College) and he became actively involved in educational work

(Resnick & Hall, 1998). Considered the "father of scientific measurement" in education, E. L.

Thorndike fostered the development and dominance of the objective achievement tests, curriculum of social efficiency and associationist learning theory.

In 1904, Thorndike and his students published some of the first standardized achievement tests in handwriting and arithmetic reasoning (Gerberich, 1963). Around World War I, multiple achievement tests were produced and commercially marketed across the United States. The achievement tests were not completely used to diagnose student learning problems. Their main purpose was to permit comparisons of class and student performance in an effort to study education in a more scientific manner (Resnick, 1980; Shepard, 2000b). For instance, the need to remedy inconsistencies in teachers' grading practices became important because it was thought to be highly inefficient. In a classic study that illustrated this point, Starch and Elliot (1913) distributed the same geometry paper to 116 high school mathematics teachers and received grades ranging from 28 to 92. Similarly, other assessments were created to find out why children were leaving school at eleven and twelve years old. In economic terms, twelve-year-old dropouts represented a "waste of scarce resources" (Cuban, 2004, p. 19).

Along with achievement testing, Thorndike also promoted a curriculum that focused on acquiring one skill at a time (Shepard, 2000a; Gerberich, 1963). Shepard (2000b) explained the movement:

> This meant taking Taylor's example of a detailed analysis of the movements performed by expert bricklayers and applying similar analyses to every vocation for which students were being prepared. Every step would be taught specifically and precise standards of measurement would be needed to assess every step of the learning process. Since it was not possible to teach every student the skills of every vocation, scientific

measures of ability were also needed to predict one's future role in life and thereby determine who was best suited for each vocation (p. 6).

Thorndike's view was that a core curriculum, concentrated in a few academic disciplines, made no sense for schools, especially at the secondary level where students were close to adulthood. Instead schools needed a vastly expanded array of curriculum options, differentiated both by student abilities and by projected future occupation and focused on the specific knowledge and skills that the student needed. Essentially, education became vocational training (Larabee, 2005).

Thorndike developed the associationist learning theory - which was a forerunner of Skinner's behaviorist learning theory (Shepard, 2000a). According to associationists, knowledge consists of a "collection of bonds: links between pairs of mental entities or between external stimulus and an internal mental response" (Resnick & Hall, 1998, p. 93). Learning involves strengthening the correct bond and decreasing the incorrect bonds. Bonds are strengthened through rewards and weakened through punishment or absence of rewards (Resnick & Hall).

The instructional practices that grew out of associationism called for instruction to be broken into singular minute steps, breaking educational processes into many steps that were reinforced through punishers or rewards. It also called for frequent testing, in which tests were made of many separate items of information based on the notion of knowledge as multiple bonds (Resnick & Hall, 1998). Textbooks adopted were composed of minimally connected bits of information with workbooks to support (Resnick & Hall).

John Bobbitt, a leader of the social efficiency movement, reflected the efficient beliefs of the day. He felt that the primary goal of curriculum design was the elimination of waste, and it was wasteful to teach people things they would never use. He felt that individuals should be

educated according to his or her capabilities; therefore, testing became the tool of scientific measurement because it was necessary to determine students' abilities and intelligence (Larabee, 2005; Shepard, 2000b).

Just as we still see the effects of Taylor's work in American industries, the proponents of the objective achievement tests, curriculum of social efficiency and associationist learning theory created lasting effects in education. To this day, objective achievement tests are the typical method of delivery for testing. In fact, Senk, Beckmann, and Thompson (1997) found in a study of 19 mathematics classes 77% of students' grades were determined by singular math objective tests. The curriculum of social efficiency is still seen today through vocational schools and ability tracking of students. Finally, the stimulus-response beliefs that accompany associationist theories are definitely alive and well in modern education. Teachers use workbooks and textbooks that are published to disseminate information in fragmented sections of the curriculum. The effects of this early era of American education have been extremely profound and lasting.

The Institutionalization of Testing in American Schools

The development of intelligence testing early in the century helped in the expansion of systematic testing in the United States. As the decades progressed, testing for various purposes became an institutionalized part of education.

The development of intelligence quotient (I.Q.) tests resulted in differing applications and beliefs between the United States and France. The French neurologist Alfred Binet had a strong influence in the development of the first I.Q. tests. He believed that intelligence was not a measurable trait, like height or weight, but that it could be improved through 'mental orthopedics' (Shepard, 2000b, p. 7). His research centered on the notion that intelligence was not

delivered at birth, but that it varied with the age of the individual being observed (Congress, 1992; Shepard).

In 1917, the United States army asked Stanford professor Lewis Terman to produce intelligence tests and a group intelligence scale. Terman revised Binet's I.Q. test, but with the philosophy that the I.Q. results were an exact measure of a fixed trait that couldn't be raised or lowered throughout a person's life (Congress, 1992; Shepard, 2000b). This later became the Alpha scale used by the Army to efficiently determine where to place recruits in jobs. The Alpha test was used for the normal and the Beta test was administered to the subnormal. Intelligence theorists poured over the data, proclaiming that a substantial proportion of American soldiers were "morons" (Congress). The idea to classify people in education was bolstered by this claim. After World War I, educators adopted the testing system under the guise of protecting the "slow witted from the embarrassment of failure while allowing the gifted to rise to their rightful levels of achievement" (Congress, p. 110).

Because measured differences were taken to be innate, the only way to manage the differences was through a highly differentiated curriculum and tracking. By the 1920s, tracking students by their I.Q. was firmly in place in American schools (Shepard, 2000b). Educational administrators and researchers urged that students, most of them from minority racial groups, be segregated in special classes and taught in a concrete and practical manner that would make them efficient workers (Shepard).

In 1929, the University of Iowa initiated the first major statewide testing program for students. The Iowa Test of Basic Skills and the Iowa Test of Educational Development results were used to evaluate both students and schools. By the late 1930s, Iowa tests were being made available to other states (Congress, 1992). It was during this period that multiple choice and

true/false questions were created for testing. The formats were quickly adapted to student tests and used in classroom assessments. The creators of the Iowa tests invented scoring machines that would make possible the "streamlined achievement testing of millions of students" (Congress). By the 1930s, multiple-choice tests were a part of public education.

Also in the 1930s, college admissions tests, like the Scholastic Aptitude Test (SAT), were implemented. The goal of this testing activity was to improve decision making about career choices and college entrance (Resnick, 1980). The tests quickly became national measures of school accountability, too. If the test scores were up, then the education system was considered to be doing well. If the scores dropped, then school quality was considered to be low. During this period, higher test scores became equated with more productive schools (Stiggins, 1999).

By the 1930s, the vocational influence of the social efficiency movement had faded and most students took traditional academic courses, but there were lasting effects. Differentiation of the curriculum and the segregation of studies by gender and social class were well established in the United States educational system (Larabee, 2005; Shepard, 2000b).

In the 1940s, tests and test batteries were introduced. These tests were designed to measure general educational development rather than knowledge of a specific content subject (Gerberich, 1963). After World War II, the administration of standardized test batteries on a regular basis throughout the student's school career was a common educational practice (Resnick, 1980).

The launch of the Russian satellite, Sputnik, was the catalyst for President Dwight D. Eisenhower's signing of the National Defense Education Act, Title V-A in 1958. This contributed to the expansion and the development of testing programs in many schools through the country (Gerberich, 1963; Resnick, 1980; Cuban, 2004). Many reforms resulted, including

raised graduation requirements in math and science, added programs for the gifted and advanced

placement high school courses (Cuban). Unlike the educational reforms that came after World

War II, which were mostly demographic, the post-Sputnik concerns were curricular, focusing on

what was being taught and how, rather than who was being taught. "The finger of blame was

pointed directly at the schools" (Rutherford, F., 1997, p. 2).

The Modern Assessment Movement

The sixties were the beginning of a new reform movement in educational testing

(Resnick, 1980). According to James Popham (2001), there was a change in the public's

perception of education. Stories were starting to surface about students graduating from high

school without the ability to read or write and this influenced the public's discontent. This

resulted in "mutterings" about requiring testing to ensure that learning occurred (p. 4-7). The

district-wide standardized tests that we see today were created during the 1960s. The tests were

commercially developed and norm-referenced in an effort to achieve local accountability

(Stiggins, 1999). At the federal level, the Economic Opportunity Act, which established Head

Start, and the Elementary and Secondary Educational Act of 1965, or Title I (Resnick, 1980)

were passed in an attempt to address gaps in educational attainment (Guskey, 2005). These laws

were required to be monitored and standardized achievement tests became the primary means for

doing so (Koretz, 2002).

In the 1970s, accountability measures expanded to a statewide level. During the seventies

and early eighties, the minimum competency testing movement spread rapidly (Linn, 2000;

Resnick, 1980). The decade began with three statewide assessments and ended with nearly 40

(Koretz, 2002; Stiggins, 1999). Minimum competency tests were most often relatively easy

multiple-choice tests used as a requirement for high school graduation (Koretz). The tests were established due to legislation that was passed with the objective of supplying parents with a "limited warranty" that a child who passed a competency test had at least mastered the fairly modest set of basic skills these tests measured. In addition, the policymakers who installed the competency tests weren't focusing on students; they were actually displaying their doubts about public school educators (Popham, 2001, p. 12).

The1980s brought in a transition to national testing (Stiggins, 1999). A Nation at Risk (National Commission on Excellence in Education, 1983) suggested that American students were outperformed on international academic tests by students from other industrial societies (Thattai, 2001). As Koretz (2002) attested, "The growing use of standardized tests for accountability was the core component of educational reform movement of the decade. New state-mandated tests were implemented, some tests were made more difficult, and the consequences of scores for students, educators, and administrators were often increased" (p. 760). Near the end of the decade, support for testing by researchers waned because of a growing awareness that coaching for these tests inflated scores and instruction was centered on the tests (Koretz, 2002; Shepard, 2000a), however, testing remained in place in educational systems.

During the 1990s, the United States became deeply involved in international testing programs (Stiggins, 1999; Stiggins, 2002). The media perpetuated the need by reporting math and science results in comparison with other countries' scores. Concerns about the results led to worries that educators "had better raise our standing among the nations of the world or risk social and economic decay" (Stiggins, 1999, p. 1). These concerns were addressed at President George H.W. Bush's educational summit, America 2000, attended by all of the nation's governors. The program led to the passing of the GOALS 2000: Educate America Act in 1994 under President

Bill Clinton (GOALS 2000). A federal presence was now established in educational assessment in the United States.

The 2000s could be called the "learning through standards and accountability" era in American education (Sloane & Kelly, 2003). This idea was formalized when President George W. Bush signed into law the No Child Left Behind (NCLB) Act (2002). This measure mandated standardized testing of every student in the United States in math and reading (Karp, 2004). The significance of standardized tests led many to believe that the United States was "once again revealing our faith in assessment of learning as a school improvement tool" (Chappuis, et al., 2004, p. 17).

In short, it is evident that some form of assessment has always existed in public education. Teachers in their classrooms have always used tests to measure achievement, ranging from oral examinations to portfolio assessments, as part of the process of student assessment. At the same time, standardized examinations have been used since the early 1800s to keep the public informed about the general quality of schools and schooling. Throughout the history of United States education, the majority of the focus has been on the latter. Stiggins (2002) urged, "If we wish to maximize student achievement in the U.S., we must pay far greater attention to the improvement of classroom assessment." (p. 758). Also, the influence of educational theorists from the early 1900s, like Thorndike, has had lasting and deeply engrained effects in the American educational system.

<center>Contemporary Conceptions about Assessment</center>

This section of the literature review is organized around a holistic notion of assessments. First, general assessments are discussed, then in an attempt to categorize, formative, summative, and official assessments are defined. The next section will include major studies and literature

reviews on the impact of classroom assessments and student achievement. Finally, teachers'

assessment literacy will be defined and explored.

Assessment in Education

Using Airasian's (2000) descriptions, there are three different types of assessments:

formative, summative and official. Formative assessments are used to change or improve

ongoing classroom processes while learning is still in progress (p. 94). Summative assessments

are used to evaluate the outcomes of instruction and take the form of tests, projects, term papers,

and final exams. Formative and summative assessments take place in the classroom; therefore,

they are also called classroom assessments. Official assessments are formal and systematic tests

that are required by the school bureaucracy for purposes such as pupil testing, grading, and

placement (p. 95).

Lambert and Lines (2000) defined assessment as "the process of gathering, interpreting,

recording, and using information about pupils' responses to educational tasks" (p. 4). They

suggested that the four purposes of assessment are: a) to provide feedback to teachers and

students about progress to support future learning, b) to provide information about the level of

pupils' achievement at points during and at the end of school, c) to provide the means for

selecting by qualification, and d) to contribute to the information on which judgments are made

concerning the effectiveness or quality of individuals and institutions in the system as a whole.

The purposes for assessing vary considerably across many groups of people within the

educational community. Nagy (2000) proclaimed the three roles of assessment as gatekeeping,

accountability, and instructional diagnosis. For example, policymakers use assessments to

monitor the quality of education and to formulate policies. Administrators and principals identify

program strengths and weaknesses to plan and improve programs. Teachers use assessments to

perform individual diagnosis, monitor student progress, carry out curriculum evaluation, and determine grades. Finally, parents and students use assessments to assess student strengths and weaknesses, determine school accountability, and make informed educational and career decisions (NCREL, 1991).

Assessing for different purposes and for different groups of people can result in intense stress for all involved. Barksdale-Ladd and Thomas (2000) interviewed 59 teachers and 20 parents in two large states that have standards, benchmarks, and standardized tests to assess students on the standards. They found that teachers and parents were unanimous about the intense stress, the undermining of meaningful instruction and learning, and the high stakes involved. Interestingly, teachers did not want to see a total dismissal of assessment practices. They encouraged best practices in assessment, including: a) providing feedback to help students improve their learning; b) making assessment a part of a student's work, which can go into a working portfolio; c) providing flexibility without dominating curriculum; d) using data to inform instruction to help teachers improve instruction; and e) using more than one type of measurement for assessing students' learning (Barksdale-Ladd & Thomas, p. 395). Additionally, Shellard (2005) suggested that teachers use frequent assessments of student performance. The data produced from these assessments can be used to determine how well students are doing and identify areas where intervention or changes in instruction are needed.

To summarize, assessments are a judgment of a pupil's achievements. There are different purposes and types of assessments, which are used for varying reasons to inform different groups of people. Although assessments can cause stress to students, teachers, parents, and all others involved, teachers conveyed the benefits of assessment when administered using best practices.

Classroom Assessment and Student Achievement

13

While many studies have focused on large-scale standardized testing programs, classroom assessments, both formative and summative, have received increased attention (Popham, 2002; Airasian, 2000, 1991; Black & Wiliam, 1998a; Stiggins, 2001, 1997; Stiggins & Conklin, 1992; Crooks, 1988). Students spend vastly greater amounts of time engaged in classroom assessment activities than in standardized testing; therefore, a greater impact is made through classroom assessment (Crooks, 1988; Stiggins & Bridgeford, 1985). In fact, teachers spend at least one third of their professional time on assessment activities that inform a wide variety of decisions made daily and those decisions directly influence students' learning experiences (Stiggins & Conklin, 1992). Reviews of the research by Natriello (1987), Crooks (1988) and, more recently, Black and Wiliam (1998b) have demonstrated that substantial learning gains are possible when teachers introduce classroom assessment strategies, particularly formative assessment strategies.

Natriello (1987) provided a conceptual framework in which the steps of classroom assessment process were broken into eight steps: a) establishing purpose for evaluating students; b) assigning tasks to students; c) setting criteria for student performance; d) setting standards for student performance; e) sampling information on student performance; f) appraising student performance; g) providing feedback to student performers; and h) monitoring outcomes of the evaluation of students (p. 156). Each of the stages of the model suggested features that teachers must attend to because they may have an impact on students. Perhaps his most significant point was that the key purposes for assessment are conflated due to the multiple uses by various stakeholders. Until educators can come to a single purpose for assessing students, then the research will continue to be irrelevant (Natriello).

Crooks' (1988) literature review from 14 specific fields of research focused on relationships between classroom assessment practices and student outcomes. The primary conclusion was that classroom assessment has powerful direct and indirect impacts. For instance, a classroom assessment guides students' judgment about what is important to learn, while it also affects students' motivation and self-perceptions of competence. Crooks also concluded that grades, considered a summative function of assessment, have been too dominant and that more emphasis should be given to using classroom assessments formatively to assist in learning. The feedback given to students should focus on the task, should be given regularly and while still relevant, and should be specific to the learning task.

Black and Wiliam (1998b) used Natriello's and Crooks reviews of the literature as a baseline for their seminal review in an effort to build on previous work, as well as include more current literature to produce a review of 250 publications. In their synthesis of studies on classroom assessment, typical effect sizes of formative classroom assessment experiments were between 0.4 and 0.7 with larger gains made in low-achieving students. They stressed that no reform or policy aimed at increasing student achievement through official testing will be successful because "learning is driven by what teachers and pupils do in classrooms" (Black & Wiliam, 1998a, p. 140). Suggestions for classroom assessment practices included enhancing feedback, actively involving students in their learning, adjusting instruction and re-teaching, and engaging students in self and peer-assessment activities. Another key point of the review revealed what they called a "poverty of practice" (p. 141) on the part of teachers. The primary difficulties teachers had with classroom assessments centered on three issues – effective learning, the negative impact of assessments, and the managerial role of assessments. Effective learning from assessments diminished because teachers encouraged rote and superficial learning

and delivery practices, such as questioning and classroom discussions have not been critically reviewed in relation to classroom assessments. Assessments typically left a negative impact due to the approaches used by teachers. They reiterated Crooks (1988) thoughts on grading; stating that it was overemphasized, while useful feedback to students was underemphasized. In addition, teachers tended to create a competitive atmosphere through comparing students with one another causing low-achieving students to believe that they cannot learn. Teachers used assessments to fulfill a managerial role and to fill a grade book, rather than to diagnose student learning. Finally, classroom assessments typically looked like official standardized tests and the results are not addressed once they were recorded.

The reviews of literature were beneficial in synthesizing the characteristics of classroom assessments. However, numerous studies, both qualitative and quantitative, have indicated that classroom assessment practices have a tremendous impact on student achievement.

In an earlier study, Benjamin Bloom and his students (1984) set out to determine what teaching style would net student achievement gains as highly as one-on-one tutoring. The control class consisted of 30 students with one teacher who taught the content and tested periodically with percentage scores given per test. The first experimental group consisted of the same classroom demographics and testing procedures, however, after tests were administered, feedback was provided followed by corrective procedures and parallel formative tests to determine the extent to which the students had mastered the subject matter. A final experimental group consisted of one student per one tutor that provided instruction that was followed by periodic formative tests, feedback and corrective procedures, and parallel formative tests. In a final achievement measure, the tutored group scored 98% higher than the control group, but

more importantly, the first experimental group that consisted of 30 students that received feedback and corrective procedures scored 84% higher than the control group.

Brookhart (1997) developed a framework that measures classroom assessment and student achievement, but also includes student self-efficacy and student effort. She tested this framework in two third grade classrooms in a small urban district (Brookhart & DeVoge, 1999). Significant relationships were found among perceptions of task, self-efficacy, effort, and achievement. This study demonstrated that the use of classroom assessment strategies will not only increase student achievement, but also students' self-perceptions about learning.

Rodriguez (2004) used Brookhart's (1997) framework to study the interrelationships of teacher assessment practices, student self-efficacy, student effort, and achievement performance. He used the United States portion of the Third International Math and Science Study (TIMSS) to estimate the relationships. At the classroom level, teacher assessment practices had significant relationships to classroom performance. He reported gains of over one and a half standard deviations on math performance arising from the effective management of classroom assessment. Comparable to Bloom's (1984) findings, Rodriguez predicted that expected achievement score gains will rival in their impact on student achievement the implementation of one-to-one tutoring instruction, with the largest gains being realized by the lowest achievers, thus reducing achievement gaps.

Similar to Rodriguez (2004), Meisel, Atkins-Burnett, Xue, and Bickel (2003) found gains made by low achievers. In their study that utilized data collected over three years, Meisel, et al. focused on curriculum-embedded classroom assessments and its impact on scores on the Iowa Tests of Basic Skills. The study took place in Pittsburgh schools in which the clientele was 71% African American, 90% received free or reduced lunch and had 9.8% mobility. The comparison

group was chosen to match schools as closely as possible on race, income, mobility, school size, and number of parents in the home. Students that were in classrooms that used curriculum-embedded classroom assessments displayed significant growth in reading and mathematics. Perhaps even more important, results of above and below average students were examined separately and they were able to demonstrate growth in both groups and across the entire study group.

In another study on classroom assessment and student achievement, teacher training was emphasized. Wiliam, Lee, Harrison, and Black (2004) wanted to determine if increased classroom assessment strategies improved student achievement on official assessments, in this case, England's school-leaving examination (the General Certificate of Secondary Education, or GCSE). Rather than simply direct teachers to use classroom assessment strategies, the researchers established training for 24 teachers over a six month span of time in exploring and planning their approach to classroom assessments. Then the teachers put the plans into action with selected classes. Results indicated that improvements equivalent to approximately one-half of a GCSE grade per student per subject were achieved.

Researchers have presented clear evidence that there is a positive relationship between teachers using classroom assessment strategies and the impact on student achievement through various measures. Knowledge and proficiency are important among classroom teachers. This concept is generally referred to as assessment literacy.

Teachers' Assessment Literacy

An increase in recent literature has called for teachers to become assessment literate. Popham (2004) referred to a lack of assessment literacy as "professional suicide" (p.82) and proclaimed that teachers are obligated to invest time toward gaining knowledge in this area.

Assessment literacy refers to a teacher's knowledge about the basic principles of sound

assessment practice, including terminology, the development and use of assessment

methodologies and techniques, and a familiarity with standards of quality in assessment.

Increasingly, this knowledge includes familiarity with alternatives to traditional measurements of

learning (Hearne, 2001).

Assessment formats

Assessment literate teachers choose appropriate formats to assess different achievement

targets and clearly understand the strengths and weaknesses of each of these formats (Stiggins &

Conklin, 1992). Examples of teacher-created assessments are teacher-developed tests and

quizzes, text-embedded tests and quizzes, classroom discussions, questioning, homework, and

seatwork. Additionally, assessment literate teachers match items with course objectives and

instruction to ensure content validity (Airasian, 1991; Black and Wiliam, 1998b). Also, teachers

provide opportunities for students to express their understanding through authentic assessments,

because they know this will ensure the interaction necessary for learning to occur (Black and

Wiliam, 1998a).

Stiggins and Conklin (1992) administered questionnaires to teachers about their levels of

concern about quality of teacher-made assessments. They found that the quality of classroom

assessments varied with grade levels and slightly with subject areas. There was an increased

concern among teachers about the improvement of teacher-made objective tests at higher-grade

levels. Also, math and science teachers were more concerned about the quality of the tests they

produced than were writing teachers.

Questions

Another suggestion for teacher-made assessments is to tap into higher-order thinking skills only after teachers have presented a solid base knowledge of what they are teaching. Stiggins (1997) originally encouraged teachers to use higher order questions in classroom assessment. For example, in a study on measuring thinking skills in classroom assessment, Stiggins, Griswold, and Wikelund (1989) analyzed writing assessments written by a group of 36 teachers from grades two through 12. Utilizing Quellmalz Taxonomy (Stiggins, Rubel & Quellmalz, 1986) – recall, analysis, comparison, inference and evaluation - the researchers found that across the grades questions of recall dominated in classroom assessments, while comparison and evaluation questions were rare. In a more recent writing, Stiggins (2001) declared that the foundation of academic competence rests on knowledge and understanding. Teachers cannot write higher-level questions without establishing a foundational knowledge base.

Classroom assessment design

Teachers create a majority of their assessments, or they adapt assessments from a text-based format. It is imperative that they have the knowledge to design, develop, use and value methods for assessing students. Teachers must also know how to match teaching objectives with assessment items, as well as use the appropriate assessment format (Black & Wiliam, 1998b). When assessing, teachers must be sure to establish a solid knowledge base but it is also important to push students into higher-order thinking. An easy way to accomplish higher levels of thinking is through performance assessments (Black & Wiliam, 1998a).

Perceptions about classroom assessments

While there are many of resources that identify strategies for improving assessment literacy, ample research has illustrated weaknesses among practicing teachers in this area. Perhaps the most telling in the lack of assessment literacy in practicing teachers were studies that

focused on teachers' perceptions of classroom assessment. Brown (2004) administered a 50-item survey to 525 New Zealand primary school teachers and principals. The four factors on the survey were improvement of teaching and learning, school accountability, student accountability and treating assessment as irrelevant. The participants agreed with the improvement conceptions and the school accountability conception, while they rejected the view that assessment was irrelevant. However, the respondents disagreed that assessment was for student accountability. Interestingly, no statistically significant differences were found in mean scale scores for each conception regardless of teacher demographics, like age, gender, or role, or for school demographics, like size, location, or socio-economic variables.

Another study of teachers' perceptions about assessment revealed teachers' underlying beliefs about assessment. The findings may help explain the struggles in changing assessment practices. Hargreaves (2005) surveyed 83 teachers asking them for their definitions of "assessment" and "learning". Through coding, the two categories that emerged from the definitions of "assessment" were assessment-as-measurement, meaning assessment used as a final judgment of students' work, and assessment-as-inquiry, meaning assessment was used to help students discover what they know about content. The first category, assessment-as-measure, was the prevalent definition by teachers in the study. Similarly, two related conceptions about "learning" were learning-as-obtaining-objectives and learning-as-construction-of-knowledge. The first conception, learning-as-obtaining-objectives, was predominant, once again demonstrating teachers' beliefs that assessments should be used as an end-product only.

Professional Development

In addition to teachers' beliefs about assessment, a lack of professional development and training may explain teachers' lack of assessment literacy. For instance, Plake, et al. (1993)

developed a survey based on the Standards for Teacher Competence in Educational Assessment of Students (AFT, NCME, NEA, 1990). The standards addressed seven broad areas in classroom assessment:
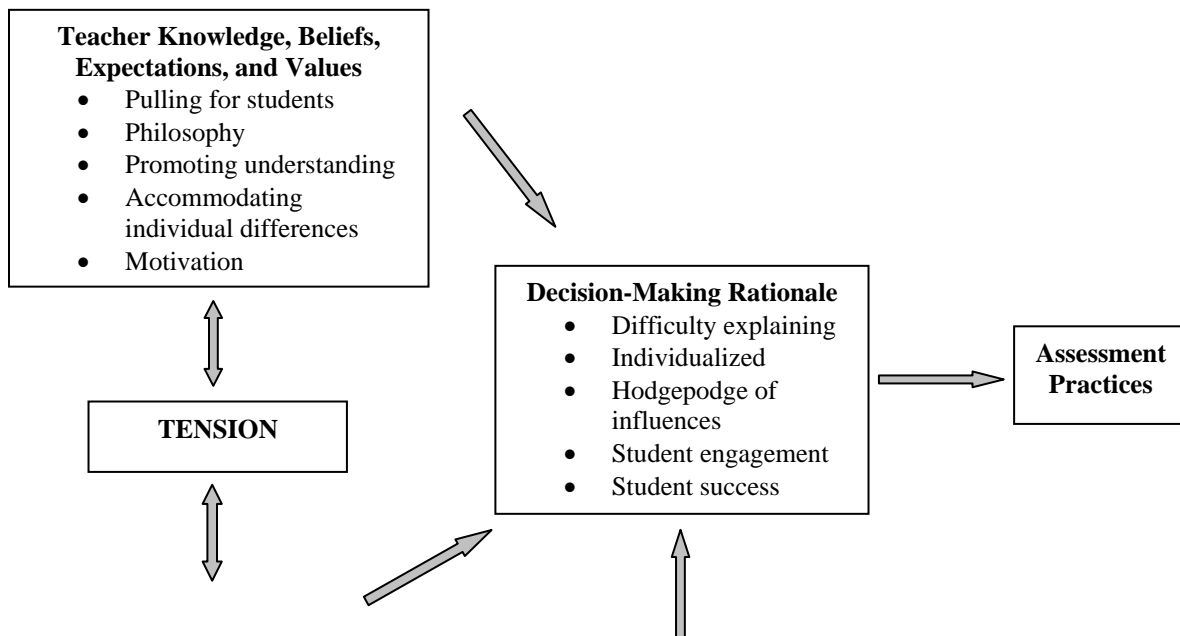
- Choosing assessment methods appropriate for instructional decisions

- Developing assessment methods appropriate for instructional decisions

- Administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods

- Using assessment results when making decisions about individual students, planning teaching, developing curriculum, and improving schools

- Developing valid pupil grading procedures

- Communicating assessment results to students, parents, other lay audiences, and other educators

- Recognizing unethical, illegal, and other inappropriate methods and uses of assessment information.

Plake, et al. (1993) administered the survey to a national sample of teachers and administrators, representing large and small school districts in rural, suburban, and urban areas. Returns were received from 42 out of the 50 United States. Overall, the highest performance area was on the subscale measuring teacher knowledge in the areas of administering, scoring, and interpreting test results. The poorest performance was on items measuring the teachers' knowledge about communicating test results.

Building on Plake's (1993) work, Metler (2004) developed the Classroom Assessment Literacy Inventory. The survey was administered to 67 undergraduate secondary education students and 101 in-service teachers in an attempt to measure teachers' assessment literacy. In-

service teachers scored highest on administering, scoring, and interpreting results of assessment and lowest on developing valid grading procedures. Pre-service teachers scored higher on choosing appropriate assessment methods and lowest, like the in-service teachers, on developing valid grading procedures.

Numerous studies focused on the severe lack of training that teachers received, both as pre-service teachers and as practicing in-service teachers (Black & Wiliam, 1998b; Cizek & Fitzgerald, 1996; Schafer, 1993). In addition, studies revealed that teachers who had taught for several years still lacked in assessment practices, thus teachers do not tend to acquire skills in this area through "on the job" training (Cizek & Fitzgerald, p. 170). Nevertheless, Cizek and Fitzgerald also found that teachers believed that it was important to do whatever would help students succeed. McMillan (2003) agreed with Cizek and Fitzgerald, but demonstrated that external factors - like accountability testing and district policies, and classroom realities - like absenteeism and disruptive behavior, created tensions that influenced teachers' decision-making in their assessment practices. The author's illustration displays this notion in Figure 3.

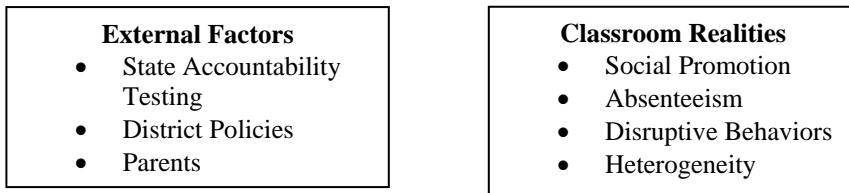| External Factors	Classroom Realities
• State Accountability Testing • District Policies • Parents	• Social Promotion • Absenteeism • Disruptive Behaviors • Heterogeneity

FIGURE 3. *Teachers' classroom assessment decision making.*
(McMillan, 2003, p. 36)

Recent studies have described training for teachers with indications of success. For instance, Lukin et al. (2004) launched reform efforts in teacher training in assessment through the University of Nebraska at Lincoln and the Lincoln Public School District. Two programs were designed to provide training to experienced teachers and two programs were developed for pre-service teachers. Evidence suggested that all of the training models had positive impacts on teacher confidence, knowledge, and skills in key areas of assessment.

Teachers' assessment literacy is a vital key in classroom assessment practices. Teachers with strong assessment literacy possess a working knowledge of when and how to design, develop, use and value a wide variety of methods for assessing student achievement (Stiggins & Conklin, 1992). A majority of the research illustrated that teachers lacked assessment literacy skills, however, there is growth in teacher training and professional development in this area.

Teacher Strategies for Classroom Assessments

Based on the extensive literature reviews (Crooks, 1988; Natriello, 1987; Black & Wiliam, 1998b) and the studies on classroom assessment and student achievement (Rogriguez, 2004; Wiliam, et al., 2004; Meisel, et al., 2003; Bloom, 1984), numerous strategies are listed as non-negotiable in classroom assessment. In this review, the strategies are broadly addressed as (1) assessment strategies used during instruction, (2) teacher adaptation of instruction per

assessment data, (3) teacher feedback per assessment data, (4) teacher development of peer and

self-assessment strategies, and (5) the formative use of summative assessments.

Assessment strategies used during instruction

Teachers should think of assessment as a process rather than an end product.

Assessments occur throughout lessons. Talk and questioning are productive forms of assessing

during instruction (Black, 1998). Talk has potential to transform learning (Gilles, 1995). Through

talk, students can give the reasoning behind strategies, tactics, and interpretations that may not be

clear from the final product. The Appalachia Educational Laboratory (AEL) (2005) developed a

framework that presents questioning as a five-stage process: a) question preparation; b)

presentation of questions; c) prompting; d) processing of student responses; and e) reflection on

questioning practice. A study was conducted testing the framework in 28 fifth- and sixth-grade

classrooms. Teachers were trained in questioning strategies and then videotaped leading a class

discussion. Questions were coded based on the framework criteria. Growth in teachers'

questioning strategies was shown in all areas.

Wolf, et al. (2006) examined the relationship between the quality of classroom talk and

academic rigor in reading comprehension lessons. The data from the study included 21 reading

comprehension lessons in several elementary and middle schools from three urban school

districts. Quantitative analyses revealed that students' ratings on talk providing knowledge and

thinking had strong, positive relationships with academic rigor. Similarly, Myers' (2005) action

research study in a kindergarten classroom showed that questioning strategies improved

comprehension in very young nonreaders. In a three-month study using interviews, rubrics and

anecdotal records, findings revealed that even younger students were able to reflect on their learning and to self-monitor their comprehension through classroom discussions.

While questioning and talk are good methods for developing comprehension during classroom assessments, it is vital that teachers record results during the assessment (Stiggins & Bridgeford, 1985). Examples for recording can be checklists, performance rating scales, or anecdotal records (Stiggins, 1997).

"One of the challenges in teaching is designing, and to be a good designer you have to think about what you're trying to accomplish and craft a combination of the content and the instructional methods, but also the assessment" (Wiggins, 2002). Performance assessments are an effective form of classroom assessment that requires good design. A performance assessment can take the form of many activities, but it is defined as the observation and rating of student behavior and products in contexts where students actually demonstrate proficiency (Stiggins & Bridgeford, 1985). Typically, students show the teacher what they know through an activity of some sort. According to Black (1998), one unifying idea among different types of performance assessments is that the activities are "direct models of reality to be assessed rather than disconnected fragments or surrogates" (p. 87). It is generally agreed that performance assessments must be designed well in advance with observable and clearly defined performance tasks (Black; Stiggins, 1997; Airasian, 1991).

Shepard, et al. (1996) wanted to study if performance assessments directly impacted student achievement. In a yearlong project, thirteen third-grade classrooms used performance assessments as a part of instruction in reading and math. Control classrooms were used for comparison. Interestingly, the achievement scores in the focus classes were not higher than the control classes, however gains were found in lower-achieving students. Sheperd, et al. pointed

out that teacher training and professional development was necessary and the lack of training influenced the study.

Detailed scoring protocols must be provided before the assessment. In a detailed review of literature, Soles (2001) presented the benefits of utilizing scoring guides throughout lessons. By sharing a scoring guide at the beginning of lessons, anxiety about success is reduced and a relationship is established between teaching and assessing. Clear, shared scoring guides reduced confusion about expectations of success. Finally, scoring guides empower students because they encouraged students to become active participants in the learning.

Assessing throughout lessons helps the teacher get to know students on an individual level. Talk and questioning provide a means for obtain assessment data. Performance assessments are a powerful type of assessment in which data are collected throughout the task. Lastly, scoring guides provide a powerful means for students to become a part of the assessment process.

Teacher Adaptation of Instruction per Assessment Data

As Black and Wiliam (1998) attest, the key part of assessment is the formative part, in which assessments are used to change or improve ongoing classroom processes while learning is still in progress (Airasian, 2000, p. 94). Therefore, teachers must use classroom assessment results to inform their teaching. If students do not exhibit success, re-teaching and possibly adjusting instruction are necessary to complete the classroom assessment cycle.

A study by Bergan, et al. (1991) on early acquisition of basic skills illustrated that re-teaching is a necessary step in cognitive development. The study involved 838 students drawn mainly from disadvantaged home backgrounds in six different regions of the United States. The teachers in the experimental group were trained to implement a measurement and planning

system which required an initial assessment input to inform their teaching at the individual level. Progress was checked after two weeks, and a new assessment was given and new decisions about students' needs were made based on the results of the assessment. The cycle continued for eight weeks. The teachers mainly used observations of skills to assess progress, and worked with open-style activities which enabled them to differentiate tasks within each activity in order to match to the needs of the individual child. The experimental group achieved significantly higher scores in reading, mathematics and science than the control group. It is also important to note that the final assessments were traditional multiple choice tests and were not adapted to match the formative teaching and learning styles of the classroom.

Teacher Feedback per Assessment Data

One of the most commonly cited necessities of classroom assessment is communicating assessment results, or feedback (Sadler, 1989; Black & Wiliam, 1998a; Cronbach, 1977; Sloane & Kelly, 2003; Bloom, B. S., Hastings, J.T., & Madaus, G., 1971). Feedback on tests, seatwork, and homework should give students guidance on how to improve, and each student must be given help and an opportunity to work on improvement. Feedback should be nonjudgmental (Brookhart, 1997) and it should focus on particular qualities of students' work, with advice on what he or she can do to improve, and should avoid comparisons with other students (Black & Wiliam, 1998a). It should be frequent and descriptive, providing students with information about their strengths as well as areas for improvement (Chappuis, Stiggins, Arter, and Chappuis, 2003).

Grading as feedback.

A tremendous amount of research has focused on the administration of grades as a form of feedback. Since grade-based decisions may have lasting academic and social consequences (Messick, 1989; Popham, 2001), teachers should weigh assessment elements according to

28

instructional emphasis (Airasian, 1991; Stiggins, Frisbie & Griswold, 1989). For example, if the main point of a writing lesson was proper nouns, then scoring should focus on proper nouns. In addition, teachers need to communicate grade criteria in advance (Stiggins, Frisbie, & Griswold).

Grades should be based on achievement factors only (Cizek, & Fitzgerald, 1996; Stiggins, Frisbie & Griswold, 1989). Brookhart (1997) found that teachers make value judgments when assigning grades. Eighty-four teachers responded to a multiple-choice questionnaire about teachers' choices when assigning grades. The multiple-choice questions were followed by an open-ended question asking why they made that particular choice. Teachers indicated that they were concerned with the consequences of grade use, especially for developing student self-esteem and good attitudes toward future schoolwork. In another study on teachers' grading practices, McMillan (2001) utilized an instrument designed to measure factors teachers consider when assigning grades. Academic achievement was the most prevalent factor, but use of extra credit, academic enablers (effort, ability, improvement, and participation) and homework and zeroes were also considered. Non-achievement factors such as effort and ability should not be incorporated into subject-matter grades (Stiggins, Frisbie & Griswold, 1989).

A negative relationship was found between feedback provided only through grades and student achievement in a study of 48 Israeli students selected from 12 classes across four schools (Butler, 1988). The students were given one of three types of feedback on assigned tasks. One-third of the students were given individually composed comments on how well their work matched evaluation criteria. A second group was given only grades, derived from the scores on the preceding session's work. The third group was given both grades and comments. In a measure of achievement, the group that received only comments increased about 33% by the second session and remained high throughout the rest of the sessions. The group that received

29

grades and comments showed significant decline across the sessions, while the group that only received grades scored low and remained that way throughout the sessions. Tests of students' interest about the task reflected the achievement scores with the group that received comments only reflecting high interest and the group that received grades only indicated low interest on the task.

Grading is an important part of classroom assessment because it is traditionally the way that teachers communicate results to students and parents (Popham, 2002). It is important to remember that grading is done appropriately by making sure that scores match the emphasis in a lesson (Airasian, 1991). It is also essential that grading criteria are presented in advance and value judgments are not considered into a score (Brookhart, 1997). Grades do not motivate all students, particularly those who feel that high grades are out of their reach (Sadler, 1989). Also, grades tend to be holistic judgments rather than descriptions of strengths and weaknesses (Cronbach, 1977; Sloane & Kelly, 2003). Through written feedback, teachers can help students "internalize quality criteria by translating them from latent to manifest and back to latent again until these criteria become so obviously taken for granted that they need no longer be stated explicitly" (Shepard, 2005, p. 68).

Feedback loop.

An important feature about feedback is that it should be used to close a gap, referring to the level in which students are performing and the desired level of achievement. If information is simply "recorded, passed to a third party who lacks either the knowledge or the power to change the outcome, or is too deeply coded (for example, a summary grade given by a teacher) to lead to appropriate action, the control loop cannot be closed, and 'dangling data' is substituted for effective feedback" (Sadler, 1989, p. 125).

Feedback is defined as having three elements: redefinition of the desired goal, evidence about the student's present position, and some understanding of a way to close the gap between the two. All must be understood to some degree before the student can take action to improve learning. (Black & Wiliam, 1998a; Sadler, 1998). According to Black and Wiliam (1998b), for an assessment to be formative the feedback provided has to be used by the student to complete the feedback loop. Students need to be trained in how to interpret feedback and how to use the feedback to improve their work (Sadler).

Using Bandura's social cognition theory, Ross and his colleagues (2002) studied how students in grades two, four, and six processed feedback. Data were collected from 71 students in Toronto, Canada, equally divided among three grades and two genders. The students reported that assessment feedback told them what they were good at, what they needed to improve on, whether they needed to work harder and clarified teachers' expectations. Virtually all of the students reported that they used evaluation data to plan future actions; however they found that older students processed feedback more thoughtfully than the younger students. They also discovered that parents' opinions about assessment feedback influenced student's thoughts on assessment and their performance.

Teachers need to be trained in writing feedback with a focus on keeping the feedback descriptive rather than evaluative (Chappius, et al., 2004). A study by Elawar and Corno (1985), focused on training teachers to focus their written feedback on deep learning rather than superficial. In this study, 18 mathematics teachers received training before working with over 500 students in three schools. A control group followed the normal practice of grading homework without comments. All students were given a pre-test and post-test. Analysis of

variance of the results showed a large effect associated with the feedback treatment, which accounted for 24% of the variance in the final achievement score.

Feedback should be more than giving a letter grade. It should be given frequently in a nonjudgmental manner. The purpose of feedback is to close the gap between where a student should be and where they actually are performing. By providing feedback, students are invited to think about their learning along with teachers. This leads to student involved classroom assessments.

Teacher Development of Peer and Self-Assessments

Involving students is an essential component of classroom assessment (Black & Wiliam, 1998a). When students become involved there is a shift from passive receptivity to active involvement in learning (Shepard, 2005; Ellis, 2001). Students are also provided the opportunity to develop skills that are essential in their professional lives, such as the ability to function as independent learners, to exercise judgment and to transfer learning from one context to another (Ellis).

Ways to involve students are student-involved record keeping, student-involved communication, and student and peer assessment (Stiggins, 2001). Some specific record keeping and communication activities are students creating visual displays of important performance criteria for bulletin boards, involving students in the process of transforming performance criteria into checklists, rating scales, and other recording methods, and having students reflect and write about their own growth over time with respect to specific criteria (Stiggins, 1997). Other examples are students helping in the development of assessment exercises, creating scoring guides, and applying the scoring guide to their own work (Stiggins, 2001).

Fortana and Ferrandes (1994) presented an example of the use of self-assessment methods and its positive impact on student achievement. They studied 25 Portuguese math teachers whom they had trained in self-assessment methods during a 20-week part-time course. They put the methods into practice with 246 students ages eight and nine and with 108 older students with ages between ten and 14. Twenty other math teachers, who were taking another course in math education, were used as the control group. Through pre- and post-testing, both groups of students made gains in math achievement, but the experimental group's mean was about twice that of the control group.

While involving students makes assessment more democratic and provides students opportunities to debate and exercise collaborative decision-making (Ellis, 2001), it has been shown to generate stress on students, particularly females. Pope (2005) looked at the effect of stress on students involved in self - and peer-assessment. Results indicated that while self- and peer-assessment increased stress, it also led to improved student performance on summative tasks.

On the other hand, Schunk (1996) asserted that providing students the opportunity to use self-assessment raised self-efficacy, skill, motivation and task orientation. Two groups of elementary students worked on the introduction of fractions. The group that used self-assessment strategies showed significant effects on the analysis of covariance (ANCOVA) in self-efficacy due to self-evaluation (13.85) at an alpha of .01 and also the interaction between goal attainment and self-evaluation (7.10) at an alpha level of .05.

True self- and peer-assessment is much deeper than asking students to trade papers to grade. Students take on a partnership role in their learning (Stiggins, 2001). Eventually, students come to understand how assessment and self-assessment affect their own academic success

(Stiggins, 2001). Higgins, et al. (1994) worked with first and second grade teachers and students in their development of assessment skills through peer evaluation. The children generated their own criteria and the quality of the criteria rose throughout the study. An example of the criteria requirements was that "it must relate to what was learned" (p. 321). This showed sophistication in the students' knowledge of the connection between learning and assessment.

Students can assess themselves only when they have a sufficiently clear picture of the targets that their learning is meant to attain (Black and Wiliam, 1998b). A study from a group of Stanford researchers presented findings that clearly illustrated this fact. Two groups of students were asked to work on group projects, but only one group was given evaluation criteria. This groups' mean on the final product was significantly higher than the group without criteria, while the group without criteria spent almost 20% of their time off-task (Cohen, et al., 2002).

Students learn how to self-assess through the teacher modeling the act of providing feedback. Then students use this knowledge to self-critique and check for those elements in their work (Shepard, 2005). Teachers have to know, or receive training in how to administer a lesson that involves self- and peer-assessment (Stiggins, 2001), but they also have to realize that student-involved assessment strategies are only a part of a larger multi-faceted assessment process that may affect student learning processes and outcomes (Marshall & Weinstein, 1984). Student-involved activities, like self- and peer-assessments, can affect the motivation and self-esteem of students (Black & Wiliam, 1998a). These activities can help increase students' control of their own welfare and will remove the mystery that too often surrounds the meaning of success in the classroom (Stiggins, 1997).

The Formative Use of Summative Data

To enhance learning, researchers have proposed using an assessment system that integrates summative and formative assessments that will improve both the quality of learning and the quality of assessments (Stiggins, 2002; Wiliam, 2000; Shepard, 2005). Wiliam claimed that improving the quality of teachers' day-to-day classroom assessment practices has a substantial enough effect to take an average country, like the United States or England, up into the top five countries on the international rankings on student achievement (p. 112).

When thinking of integration between formative and summative assessments, it is important to remember the meanings of the terms. Bloom, et al. (1971) were the first to define the terms. They defined summative evaluations as assessments given at the end of units, mid-term, and at the end of a course, which are designed to judge the extent of students' learning in a course (p. 117). They contrasted these with formative evaluations, which elicit evidence that yields interpretations that form the basis for successful action in improving performance (Black & Wiliam, 1996). To integrate formative and summative assessment practices, teachers take summative assessment data and use it to make instructional decisions to improve student performance (Black & Wiliam, 1998a).

Black and Wiliam (1996) stressed that the terms formative and summative not be applied to the assessments, but to the functions that they served. Shepard (2005) stressed that "knowledge is constructed, and that learning and development are culturally embedded, socially supported processes" (p. 66). This learning theory provides the big-picture understandings needed to change teaching and assessment practices.

An example of a school that worked to integrate formative and summative assessment was described by Many and Jakicic (2006). In the qualitative study, teachers focused on assessments, and they realized that they were inconsistent in their classroom assessment

practices with a wide range between highly summative and loosely formative assessments. They worked to develop a continuum of assessments throughout the middle school years constant to all of the teachers. They created a balanced assessment system that looked at four categories of assessments: classroom assessments, common assessments, district-level assessments, and official assessments. Each of the assessments were analyzed and discussed. Teachers discovered that the conversations about the assessments affected their teaching and what they felt was important. Additionally, nearly 90% of the school's students improved on official assessments.

In another study, Henning (2006) studied 24 elementary and middle school teachers as they analyzed standardized achievement test scores in a Teacher Leadership Program at the University of Northern Iowa. Findings indicated that teachers were adept at analyzing and disaggregating data after they were provided with training. Findings also showed that the teachers found numerous ways to analyze the data creating a wide variety on information.

Integration of summative and formative assessments is perhaps the key factor to the successful use of classroom assessments. Teachers take summative data but rather than use it to make a final judgment about student performance it is used formatively to provide students with the information they need to improve achievement. The extent to which instruction is guided by assessment data depends on the leadership of the principal. Specifically, it depends on the principal's ability to model tools and strategies for using assessment data to improve instruction; help the staff acquire the requisite skills to use assessment data for instructional decision-making; and establish a school-wide norm that instruction will change based on assessment data (Fox, 2003).

Instructional Leadership

Since the beginnings of the principalship in American education, a distinctive definition for the position never emerged. Due to the technical and bureaucratic nature of running a building, principals were originally thought of in a managerial role. However, in every decade some attention has been paid to the principal's impact on teaching and learning processes (Beck & Murphy, 1993).

Traditionally, principals were expected to be the manager of the building. During the eighties, there was a shift in attention to educational roles with both teachers and principals receiving emphasis (Murphy, 1985). A prevailing assumption of this era was that the principal should become directly involved with the teaching and learning processes of their building (Beck & Murphy, 1993). Principals were to "intervene" to ensure that teachers focused on the central mission of the school (Beck & Murphy). More than ever, principals were expected to become directly involved with instruction rather than the more traditional role as managerial leader.

Research in the nineties "pushed the principalship in new directions, focusing on the larger external forces that were shaping the future of schools" (Beck & Murphy, 1993). The focus expanded from instruction to principal as servant, organizational architect, social architect, educator, moral agent, and a person in the community (Beck and Murphy).

Later, attention on instructional leadership shifted from a focus from instructional leadership as direct involvement in classrooms to a multi-dimensional role in which the administrator fosters an environment promoting the improvement of instruction through non-managerial approaches (Macmillan, Meyer & Sherman, 2001). Administrators need to know about instructional leadership to: a) create coherence in improvement efforts; b) finely balance mandate and empowerment; 3) model learning – openness to new ideas, willingness to be driven by results, and persistence in the face of difficulty (Lashway, 2002).

Leithwood and Duke (1998) described instructional leadership as focusing on the behaviors of teachers as they engage in activities directly affecting the growth of students. At the heart of this definition is the conception that instructional leaders directly impact student achievement (Hallinger & Heck, 2000; Smith & Andrews, 1989; Andrews, Basom, & Basom, 1991). Sheppard (1996) referred to two views of instructional leadership, "narrow" and "broad." The narrow view of principal instructional leadership identifies it as a separate component of the principal's responsibility, excluding behaviors that focused on school climate and teacher development (Sheppard). The broad perspective involves all activities that affect student learning (Murphy, 1988). Taking a broad view of instructional leadership, the common themes that emerged from the literature focused on instructional leader attributes and job functions, instructional leader knowledge and teacher development.

Instructional Leader Attributes and Job Functions

While Sergiovanni (1991) suggested that research on instructional leadership attributes and functions needed to be "situationally specific" (p. 90), many researchers determined characteristics that suggest strong instructional leadership. Smith and Andrews (1989) compiled a list of characteristics from existing literature that suggest strong instructional leadership. This list includes "high energy, assertiveness, ability to assume the initiative, openness to new ideas, tolerance for ambiguity, a sense of humor, analytic ability, and a practical stance toward life" (p. 8). In addition, referent power, strong motivation, and high self-esteem were influential attributes (p. 39).

In their review of related literature, Andrews, Basom and Basom (2001) found three inherent attributes common in strong instructional leaders: vision, the ability to communicate that vision, and the ability to create trust in the workplace. Building on the notion of trust, Beatty

and Brew (2004) studied principals' work in establishing trust in their schools. Using an emotional epistemologies theoretical framework, the researchers interviewed and analyzed written responses given by 42 principals in an educational leadership program. They found that effective principals who were comfortable creating a high amount of trust established a "Vygotskian zone of proximal development" (p. 351) with their teachers when introducing instructional reform. By establishing trust, instructional leadership was more effective (Beatty and Brew).

Some discussions on the functions of an instructional leader focused on the nature of the job, stating that instructional leaders set "high academic expectations, review lesson plans, supervise classroom instruction, and monitor curriculum" (Lashway, 2003, p.4). Smith and Andrews (1989) explained instructional leadership as a combination of several tasks, such as supervision of classroom instruction, staff development, and curriculum development.

Common descriptions of the functions of instructional leaders included the necessity to provide constructive support and obtain the resources and materials necessary while keeping abreast of the latest developments in teaching, learning, motivation, classroom management, and assessment (Hoy and Hoy, 2003). For example, Smith and Andrews (1989) found in their review of the literature that principals who were instructional leaders demonstrated skills as a resource provider, instructional resource, communicator, and were a visible presence in the school.

Similarly, Hallinger and Murphy (1985) studied the instructional leadership behaviors of ten elementary principals. They administered a questionnaire, The Principal Instructional Management Rating Scale, to teachers, principals and the principals' supervisors. The measure was divided into three main functions of instructional leadership. The three functions were

defines the school mission, manages instructional programs, and promotes school culture. They

also analyzed principals' documents such as teacher evaluation reports, newsletters, and school

bulletins. They found the principals in this study frequently engaged in instructional leadership

behaviors. They also noted that the high instructional leadership behaviors might have been

influenced by the introduction of assistant elementary principals to the district. They speculated

that by adding assistant principals, principals had more time for instructional leadership

activities.

Specific leadership behaviors and functions were the focus of the Marzano, Waters, and

McNulty (2005) meta-analysis of 69 other studies conducted that focused on specific leadership

behaviors and their impact on student academic achievement. Table 1 illustrates the

responsibilities that Marzano, et al. found to influence instructional leadership practices and their

correlations with student academic achievement.

Table 1

Instructional Leadership responsibilities and their correlations (r) with student academic achievement.

Responsibility	The Extent to Which the Principal…….	Average *r*
Intellectual stimulation	Ensures faculty and staff are aware of the most current theories and practices and makes the discussion of these a regular aspect of the school's culture	.24
Involvement in Curriculum, Instruction, and Assessment	Is directly involved in the design and implementation of curriculum, instruction, and assessment practices	.20
Knowledge of Curriculum, Instruction, and Assessment	Is knowledgeable about current curriculum, instruction, and assessment practices	.25
Monitoring/	Monitors the effectiveness of school practices and	.27

Evaluating   their impact on student learning

Marzano, Waters and McNulty (2005) also indicated that a critical function of leadership
is the knowledge to "do the right work" (p. 76) in their schools. Strong instructional leaders
know the direction to lead their schools when implementing comprehensive school reforms that
lead to increased student achievement. "There is a significant difference between the existence of
programs commonly recommended for practice in middle level literature and a thorough
understanding of that literature coupled with the effective implementation of the programs"
(Valentine, Clark, Hackmann, & Petzko, 2004, p. 108). Instructional leaders must have a
working knowledge of best practices.

Instructional Leader Knowledge

Efficient, thoughtful and purposeful daily and ongoing functions of a principal are
essential for a school's success; however a principal also has to have a deep comprehension of
the theories and knowledge necessary to maintain a successful instructional environment (Fisk
and Resnick, 1999). Fullan (1997) emphasized, "when it comes to learning, effective leaders
need to be greedy" (p. 45).

In a national study, Valentine, et. al (2004) determined that leadership knowledge was a
characteristic of highly successful middle schools:

…what was impressive about the (highly successful) schools…was the degree of

knowledge and understanding that principals and teachers had about the research

and literature on best programmatic practices in the middle level. They expressed

strong support for middle level programs such as interdisciplinary teaming,

exploratory courses, and advisory opportunities; they implemented those

programs very effectively; and they articulated why they implemented the

programs and how the programs were effective for their students (p. 109).

In a case study of a problems-based instructional leader development course, Hallinger

and McCary (1990) argued that research on instructional leadership needed to address the

knowledge and thinking that underlies the exercise of leadership. Utilizing a computer

simulation program, principal trainees were provided with $30,000 and 2,000 hours of staff

energy to address the problem of low test scores. Simply investing in the most cost effective

program was not enough for success. The principal trainees found that knowledge of

instructional methods was necessary for school improvement. One participant stated, "Efforts to

use new instructional strategies are unlikely to bring about lasting changes in a teacher's

behaviors if peers and the principal do not understand or appreciate the significance of the new

strategy" (p. 100). Hallinger and McCary concluded that good instructional leadership involved

proficient planning and management. Additionally, instructional leaders must have the ability to

use forethought before acting, as well as the ability to strategically respond in regards to a

specific purpose.

Ruff and Shoho's (2005) collective case study also focused on the cognitive levels of

instructional leadership of three elementary principals at varying stages of their careers. They

utilized Brewer's definition of mental models as a means of communicating the tacit

assumptions of instructional leadership. They defined mental models as the "specific knowledge

structures that are constructed to represent a situation through the use of generic knowledge" (p.

557). The three participants were principals from highly successful urban schools in San

Antonio, Texas. The first principal was a first year principal, the second was in her sixth year,

and the third was a nationally recognized principal with over 20 years of experience. Findings

supported their beliefs that principals use different cognitive levels depending on their stage in their career. The first year principal's mental models were appropriate, but surface level. The sixth year principal focused on how involved she should become in the process of instructional leadership. They found the veteran principal had the ability to adjust and shift her perceptions of what was important to meet the needs of a particular problem. These findings supported their beliefs that principals use different cognitive levels depending on their stage in their career.

In a study of administrators' knowledge about assessments, or assessment literacy, Impara and Plake (1995) found administrators to be more knowledgeable than teachers. Strength areas for administrators included an understanding for selecting assessment strategies. In addition, they performed well on communicating assessment results, but were less proficient in the interpretation of standardized test results.

In a case study on New York City's District Two, Fisk and Resnick (1999) utilized cognitive apprentice theory in an attempt to understand the development of knowledge in principals as instructional leaders.

> Building an effective community of principals is about both things—the craft of teaching and learning and the building of strong interpersonal relationships. I believe no effective learning can go on without very strong personal relationships. But relationships can't substitute for deep knowledge. The challenge is to build those relationships around studying teaching and improving instruction for kids and a belief system about learning (Fisk and Resnick, 1999, p. 56).

Instructional leaders must possess a working knowledge of the curriculum and instructional strategies, but they are not alone. Part of an instructional leader's job is intellectual stimulation of teachers through teacher development. In fact, Marzano, et al. (2005) found a .24

correlation in importance of intellectual stimulation of teachers. This refers to the extent to which the school leader ensures that teachers and staff are presented the most current theories and practices regarding best instructional practices and creates an open environment where discussions of the practices are a regular aspect of the school's culture.

Teacher Development and Intellectual Stimulation

Teacher development is an important facet of instructional leadership (Blase and Blase, 1998). In successful middle schools, continual learning permeates every aspect of the building starting with the principal creating a culture of learning for teachers and teachers modeling learning for students through attendance of graduate school, readings, and small and whole group studies (Valentine, et. al, 2004). The principal's role is to cultivate teachers' intrinsic motivation through frequent classroom visits and attendance of team meetings to continually work to improve their teaching abilities while remaining focused on the school's vision (Jackson & Davis, 2000).

Darling-Hammond, Ancess, and Falk (1995) described the importance of leadership and teacher development in successful schools:

> Each of the schools is collaboratively managed and in each case the principal encourages a culture of inquiry that provides a context for adult engagement. School leaders have encouraged multiple opportunities for professional growth in various forms and shapes, inside and outside the school community, for teachers at all states of development. Each understands the ebb and the flow of change, respect swells of enthusiasm and the undertow of resistance, and tolerates the directional cross currents, eddies, and still waters that characterize their faculty's

engagement in change. All the while, their commitment to their vision remains steadfast (p. 265).

In a study by Leithwood, et al. (2002), findings indicated that school leaders needed to realize the potential use of official testing policies as a means to improve teaching and learning and principals must recognize such outcomes might significantly increase the likelihood of teachers using official assessment initiatives for their own purposes, rather than dismissing them out of hand. Moreover, Blase and Blase (1998) supported the idea by declaring that everyone in a school is a learner; therefore a school is comprised of a "community of learners" (p. 127). They also emphasized that it is essential that all members act as colleagues and coaches; thus they all learn from one another. Through open discussions on curriculum, instruction, philosophy, research and all things related to learning, instructional leadership occurs (Blase and Blase).

Illustrating the need for teacher development when looking at instructional leadership, Marks and Printy (2003) proposed that school leadership needed to be analyzed through a lens of "integrated leadership" (p. 377), or the combination of transformational leadership and instructional leadership. Transformational leaders motivate followers by "raising their consciousness about the importance of organizational goals and by inspiring them to transcend their own self-interest for the sake of the organization" (p. 375). The researchers wanted to thoroughly investigate instructional leadership in 24 schools that were deemed exceptional on a national level by the Center for Organization and Restructuring of Schools. They found that strong transformational leadership by the principal was essential in supporting teacher commitment; therefore, when teacher commitment increased, so did instructional leadership. Thus, they proposed that instructional leadership could be transformational. They also proposed

that active principal and teacher collaboration around instructional matters enhanced the quality of teaching and student performance.

Conceptions about instructional leadership have changed throughout the decades. Prior conceptions about principals, particularly as managers, are important to understand to fully appreciate how important the work was from the eighties, as well as current notions on instructional leadership. The demands of standardized testing have created a need for principals to become instructional leaders (Quinn, 2002).

<center>Teachers' Assessment Practices and Leadership</center>

The educational reforms initiated by No Child Left Behind have created a change in the principal's role as an instructional leader. New forms of assessment, school reform initiatives, the growing number of mandated assessments, and a host of individuals demanding accountability have converged on our schools. Dealing with these and other assessment issues requires a type of leadership that Cizek (1995) coined as "assessment leadership." Assessment leadership requires four characteristics that school and district-level administrators must possess: a) an intimate knowledge of what occurs in classrooms, b) a clear focus on the desirable educational outcomes targeted by the curricular programs in place, c) familiarity with the purposes of a given assessment and audiences to which results are to be reported, and d) some understanding of fundamental assessment concepts (Cizek, p. 247).

Considering the growing number of mandated assessments and growing number of groups interested in various assessment outcomes, the principal must have a vision regarding how the various assessment parts fit together (Cizek, 1995). When establishing the assessment vision and goals in a school, it is essential that the principal have an appreciation of the integral

relationships among teaching, learning and assessment and the resulting influence on student achievement (Noonan & Renihan, 2006; Girvin, 2005).

It is also essential to include the teachers in the assessment leadership of the school. The act of bringing together teachers to focus on assessment data lends itself to the research on professional learning communities (DuFour & Eaker, 1998). The common tenets of professional learning communities are: a) ensuring that students learn b) creating a culture of collaboration, c) focusing on results, and d) hard work and commitment (DuFour, Eaker, & DuFour, 2005). Classroom assessment data provides the stimulus for each of the tenets of a professional learning community (Stiggins, 2004).

Conclusion

The history of assessment practices in the United States revealed an institution steeped in beliefs that were primarily based on the notion that assessments were used as an end product designed as a measure of learning. Traditions in education that were established during the Industrial Revolution have had lasting impacts in current education.

The accountability movement has increased the purposes for assessment, thus, there has been amplified stress that comes with assessments due to the varying involved parties like policymakers, principals, teachers, students and parents. This review of literature explored the use of classroom assessments as a possible solution to decreasing the focus on official assessments and using assessments to increase student achievement.

The literature on current conceptions about classroom assessments indicated that they are an important part of the assessment process. Classroom assessments are not only "one of our indicators of educational outcomes, but classroom assessments are also part of the very instructional treatments that produce the desired outcomes" (Stiggins & Conklin, 1992).

Numerous literature reviews and studies have bolstered this claim. Three points are noteworthy about the existing literature. First, teacher knowledge of classroom assessment, or assessment literacy, is essential (Chappuis, Stiggins, Arter & Chappuis, 2004). Many teachers did not receive assessment training during teacher certification (Stiggins, 2001; Black & Wiliam, 1998a; Shepard, et al. 1996), so professional development is essential (Chappuis, et. al). Second, classroom assessments involve a broad range of strategies. Teachers may be involved in some activities more than in others due to the nature of the grade levels and content areas they are required to teach. The existing literature has suggested that grade levels and subject areas may account for some variations in classroom assessment (Stiggins & Conklin, 1992).

Ideas about instructional leadership have evolved throughout the years. In today's climate of accountability, there is a need for instructional leaders to transcend classroom management and evaluation. Today's instructional leaders need to work to create a community of learners in which teachers see assessment as a tool that can be used to improve overall student achievement. This focus on assessment calls for principals who develop strong "assessment leadership" (Cizek, 1995). Assessment leadership requires the principal to comprehend the "big picture" of new forms of assessment, school reform initiatives, and the growing number of mandated assessments to create a vision for the school (Cizek, p. 247). The assessment vision is the catalyst for teacher collaboration where teachers and administrators come together to study assessment data results with a focus on improving student learning (DuFour & Eaton, 1998).

References

Airasian, P. (2000). *Assessment in the classroom* (2nd ed.). Boston: McGraw-Hill.

Airasian, P. (1991). *Classroom assessment*. Boston: McGraw-Hill.

American Federation of Teachers, National Council on Measurement in Education, &
National Education Association (AFT/NCME/NEA) (1990). Standards for teacher
competence in educational assessment of students. *Educational Measurement: Issues and
Practice, 9*(4), 30-32.

Andrews, R., Basom, M., & Basom, M. (1991). Instructional leadership: Supervision that
makes a difference. *Theory into practice, 30*(2).

Appalachian Educational Laboratory (AEL), (2005). *Assessing the relationship between
questioning and understanding to improve learning and thinking and student
achievement in mathematics: A pilot study*. Retrieved November 1, 2006 from
www.edvantia.org

Arhar, J. M. (1992). Interdisciplinary teaming and social bonding of middle level
students. In J.L. Irvin (Ed.), Transforming middle level education: Perspectives
and possibilities. Needham Heights, MA: Allyn & Bacon.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning.
Educational Psychologist, 93(2), 117-149.

Barksdale-Ladd, M. & Thomas, K. (2000). What's at stake in high-stakes testing?
Teachers and parents speak out. *Journal of Teacher Education, 51*(5), 384-397.

Beatty, B.R. & Brew, C.R. (2003). Trusting relationships and emotional epistemologies:
A foundational leadership issue. *School Leadership and Management 24*(3), 329-
356.

Beck, L. & Murphy, J. (1993). *Understanding the principalship: Metaphorical themes   1920s-1990s*, New York: Teachers College Press.

Bergan, J., Sladeczek, I., Schwarz, R. & Smith, A. (1991). Effects of a measurement and planning system on kindergarteners' cognitive development and educational programming. *American Educational Research Journal, 28*, 683-714.

Black, P. (1998). *Testing: Friend or foe?* London: Routledge-Falmer.

Black, P. & Wiliam, D. (1998a). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 139-148.

Black, P. & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in Education, 5*, 7-74.

Black, P. & Wiliam, D. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 1-11.

Black, P., Harrison, C., Lee, C. Marshall, B., & Wiliam, D. (2004). Working inside the black box. *Phi Delta Kappan. 86*(1), 13-22.

Blase, J. & Blase, J. (1998). *Handbook of instructional leadership: How really good principals promote teaching and learning*, Thousand Oaks, CA: Corwin Press, Inc.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-on-one tutoring. *Educational Researcher, 13*(6), 4-16.

Bloom, B. S., Hastings, J.T., & Madaus, G. (1971). *Handbook on the formative and summative evaluation of student learning*. New York: McGraw-Hill.

Bottoms, G. (n.d.). What school principals need to know about curriculum and instruction.

Retrieved April 8, 2007 from www.sreb.org

Bowman, M. & Valentine, J. (1984). Audit of principal effectiveness: A user's manual.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in

motivating student effort and achievement. *Applied Measurement in Education, 10*(2),

161-180.

Brookhart, S. M. & DeVoge, J. G. (1999). Testing a theory about the role of classroom

assessment in student motivation and achievement. *Applied Measurement in Education,*

12(4), 409-425.

Brown, A. (1994). The advancement of learning. *Educational Researcher, 23*(8), 4-12.

Brown, G. (2004). Teachers' conceptions of assessment: Implications for policy and professional

development. *Assessment in Education, 11*(3), 301-318.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving

and ego-involving evaluation of interest and performance. *British Journal of Educational*

Psychology, 58, 1-14.

Chappuis, S., Stiggins, R.J., Arter, J., & Chappuis, J. (2004). *Assessment FOR learning*.

Portland: Assessment Training Institute.

Cizek, G. (1995). The big picture in assessment and who ought to have it. *Phi Delta Kappan,*

77(3), 246-249.

Cizek, G. & Fitzgerald, S. (1996). Teachers' assessment practices: Preparation, isolation, and the

kitchen sink. *Educational Assessment, 3*(2), 159-179.

Cohen, E., Lotan, R., Abram, P., Scarloss, B., & Schultz, S. (2002). Can groups learn? *Teachers*

College Record, 104(6), 1045-1068.

Congress of the United States. (1992). Lessons from the past: A history of educational testing in

the United States. In Office of Technology Assessment (Eds.), *Testing in American schools: Asking the right questions*. Washington, DC: U.S. Office of Technology Assessment.

Cronbach, L. J. (1977). *Educational psychology* (3rd ed.). New York: Harcourt Brace Jovanovich.

Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-480.

Cuban, L. (2004). Looking through the rearview mirror at school accountability. In K. Sirotnik (Ed.), *Holding accountability accountable: What ought to matter in public education*. New York: Teachers College Press.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York: Teachers College Press.

Darling-Hammond, L. & Wise, A. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal, 85*(2), 315-336.

Dewey, J. (1993). *How we think: A restatement of the relation of reflective thinking to the educative process.* Boston: Houghton Mifflin Company.

DuFour, R. & Eaker, R. (1998). *Professional learning communities at work: Best practicies for enhancing student achievement.* Bloomington, Indiana: National Educational Service.

DuFour, R., Eaker, R., & DuFour, R. (2005). *On common ground: The power of professional learning communities.* Bloomington, Indiana: National Educational Service.

Earl, L., Levin, B., Leithwood, K., Fullan, M., Watson, N., Torrance, N., Jantzi, D., Mascall, B., & Volante, L. (2003). England's National Literacy and Numeracy Strategies: Final report of the external evaluation of the implementation of the strategies. Department of

Education and Employment, England.

Elawar, M. & Corno, L. (1985). A factorial experiment in teachers' written feedback on student

  homework: Changing teacher behavior a little rather than a lot. *Journal of Psychology,*

  *77(*2), 162-173.

Ellis, G. (2001). Looking at ourselves – self-assessment and peer assessment: Practice examples

  from New Zealand, *Reflective Practice, 2*(3), 289-302.

Farr, R. (1992). Putting It All Together: Solving the Reading Assessment Puzzle. *Reading*

  Teacher, 46(1), 26-37.

Fienberg, H. (2002). *Davis' failed test*. Retrieved December 12, 2006 from

  www.hfienberg.com/clips/ca_test.htm

Fisk, E. & Resnick, L. (1999). *Developing principals as instructional leaders.*

  Philadelphia, Pennsylvania: University of Pittsburgh, HPLC Project, Learning

  Research and Development Center.

Fortana, D. & Fernandes, M. (1994). Improvements in mathematics performance as a

  consequence of self-assessment in Portuguese primary school pupils. *British Journal of*

  Educational Psychology, 64, 407-417.

Fox, D. (2003). Guiding instruction through assessment: What principals need to know.

  Leadership.

Fullan, M. (1997). What's worth fighting for in the principalship? New York: Teachers College

  Press.

Gerberich, J. R. (1963). The development of educational testing. *Theory into Practice,*

   2(4), 184-191.

Gilles, C. (1995). Reflections. *Primary Voices K-6, 3*(1), 39-43.

Girvin, N. (2005). *The principal's role in k-12 professional development.* Retrieved from

http://www.askasia.org

GOALS 2000: Educate America Act (1994). *The Clinton Administration Education*

Program. Retrieved September 14, 2006 from www.ed.gov/G2K/

Graue, M. (1993). Integrating theory and practice through instructional assessment.

Educational Assessment, 1(4), 283-310.

Green, L. (2001). *Practicing the art of leadership: A problem-based approach to*

implementing the ISSLC standards. Columbus, OH: Prentice-Hall.

Guskey, T. R. (2005). Formative classroom assessment and Benjamin S. Bloom: Theory,

research, and implications. *Man and Society*, *57*(4), 219-233.

Hallinger, P. & Heck, R. (2000). *Exploring the principal's contribution to school*

effectiveness, 1980-1995. Washington, D.C.: Institute for Educational

Leadership.

Hallinger, P. & McCary, C.E. (1990) Developing the strategic thinking of instructional

leaders*. The Elementary School Journal, 91*(2) 89-108.

Hallinger, P. & Murphy, J. (1985). Assessing the instructional management behavior of

principals. *The Elementary School Journal, 86*(2), 217-247.

Hallinger, P., Bickman, L. & Davis, K. (1996). School context, principal leadership, and

student reading achievement. *The Elementary School Journal, 96*(5), 527-549.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box.

Cambridge Journal of Education, 35(2), 213-224.

Hearne, J. (2001). Assessment as a tool for learning. *New Horizons for Learning*.

    Retrieved November 12, 2006, from www.newhorizons.org/strategies/assess/hearne.htm

Henning, J. (2006). Teacher leaders at work: Analyzing standardized achievement data to

    improve instruction. *Education, 126*(4), 729-737.

Herbert, G. (1997). Practical assessment and testing in a secondary school. In C. Cullingford

    (Ed.) *Assessment versus Accountability*. London: Cassell.

Higgins, et al., (1994). Placing assessment into the hands of young children: A study of student-

    generated criteria and self-assessment. *Educational Assessment, 2*(4), 309-324.

Hoy, A. & Hoy, W. (2003). *Instructional leadership.* 1st Edition. Boston, MA: Allyn and Bacon.

Impara, J.C. & Plake, B. (1995). Comparing counselors', school administrators', and teachers'

    knowledge in student assessment. *Measurement and Evaluation in Counseling and*

    Development, 28(2).

Jackson, A. & Davis, G. (2000). *Turning points 2000: Educating adolescents in the 21st century*.

    New York: Teachers College Press.

Johnson, B. & Christensen, L. (2000). *Educational research: Quantitative and qualitative*

    approaches. Needham Heights, MA: Allyn & Bacon.

Karhuse, A. (2007). House and senate have differing priorities for NCLB reauthorization,

    Newsleader, 54(5), 1, 11.

Karp, S. (2004). No child left behind: The test. Rethinking Schools Online. Retrieved December

    2, 2006, from http://www.rethinkingschools.org/archive/19_01/test.191.shtml

Koretz, D. (2002). Limitations in the use of achievement tests as measures of educators'

    productivity. *The Journal of Human Resources, 37*(4), 752-777.

Krug, S. (1992). Instructional leadership, school instructional climate, and student learning

outcomes (Tech. Rep. No. 143). Urbana, IL: University of Illinois, National Center for School Leadership

Lambert, D. & Lines, D. (2000). *Understanding assessment*. London: Routledge/Palmer.

Larabee, D. (2005). Progressivism, schools and schools of education: An American romance, *Paedagogica Historica, 41(1&2),* 275-288.

Lashway, L. (2002). Developing instructional leaders. *ERIC Digest, 160.*

Le, V., Stecher, B., Lockwood, J., Hamilton, L. & Robyn, A. (2006). Improving mathematics and science education: A longitudinal investigation of the relationship between reform-oriented instruction and student achievement (ED494015). Washington, DC: RAND Corporation.

Leithwood, K. & Duke, D. (1998). Mapping the conceptual terrain of leadership: A critical point of departure for cross-cultural studies. *Peabody Journal of Education,* 73, 31-50.

Leithwood, K., Steinbach, R. & Jantzi, D. (2002). School leadership and teachers' motivation to implement accountability policies. *Educational Administration Quarterly, 38*(1), 94-119.

Linn, R. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4-16.

Lukin, L., et al. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 26-36.

Macmillan, R., Meyer, M., & Sherman, A. (2001). *The evolving role of educational administrators.* Paper presented for the 2001 Pan-Canadian Education Research Agenda Symposium Teacher Education/Education Training: Current Trends and

Future Suggestions.

Many, T. & Jakicic, C. (2006). A steadily flowing stream of information gives teachers much-needed data. *Journal of Staff Development, 27*(1), 46-51.

Marks, H. & Printy, S. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly, 39*(3), 370-397.

Marshall, H. & Weinstein, R. (1984). Classroom factors affecting students' self-evaluations: An interactional model. *Review of Educational Research, 54* (3), 301-325.

Marzano, R., Waters, T., & McNulty, B. (2005). *School leadership that works: From research to results.* Aurora, CO: Mid-Continent Research for Education and Learning

McMillan, J. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 34-43.

McMillan, J.H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice,* 125-136.

McTighe, J. & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership, 63*(3), 10-17.

Meisels, S., Atkins-Burnett, S., Xue, Y, & Bickel, D. (2003). Creating a system of accountability: The impact of instructional assessment on elementary children's achievement test scores. *Education Policy Analysis Archives, 11*(9).

Mertler, C. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education, 33*(1), 49-64.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment.

Educational Researcher, 18(2), 5-11.

Missouri Department of Elementary and Secondary Education (2006). *Missouri School Directory*. Retrieved January 1, 2007, from http://dese.mo.gov/directory/download.html

Missouri Department of Elementary and Secondary Education (2007). *2006 Schools in school improvement and restructuring.* Retrieved April 10, 2007, from http://dese.mo.gov/divimprove/nclb

Morgan, G. (2006). *Images of Organization.* Thousand Oaks, CA: Sage Publications.

Murphy, J. (1988). The instructional leadership role of the school principal: An analysis. *Educational evaluation and policy analysis, 10*(2), 71-79.

Murphy, J. (1985). *Effective schools: What the research reveals.* APEX: Case Report, 1(5), 1-4.

Myers, P. (2005). The princess storyteller, Clara Clarifier, Quincy Questioner, and the wizard: Reciprocal teaching adapted for kindergarten students. *The Reading Teacher, 59*(4), 314-324.

Nagy, P. (2000). The three roles of assessment: Gatekeeping, accountability, and instructional diagnosis. *Canadian Journal of Education, 25*(4), 262-279.

National Commission on Excellence in Education (1983). *A nation at risk*. Retrieved July 12, 2006 from www.ed.gov/pubs/NatAtRisk/index/html

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist, 22*, 155-175.

No Child Left Behind (2002). Retrieved November 22, 2005 from http://www.ed.gov/nclb/landing.jhtml

Noonan, B. & Renihan, P. (2006). Demystifying assessment leadership. *Canadian Journal of Educational Administration and Policy, 56.*

North Central Regional Educational Laboratory (NCREL) (1991). *What does research say about assessment?* Retrieved April 27, 2006, from www.ncrel.org/sdrs/areas/stw_esys/4assess.htm

Pederson, S. & Liu, M. (2003). Teachers' beliefs about issues in the implementation of a student-centered learning environment. *Educational Technology Research and Development, 51*(2), 57-76.

Plake, B., Impara, J. & Fager, J. (1993). Assessment competencies of teachers: A national survey, *Educational Measurement: Issues and Practice,* 10-39.

Pope, N. (2005). The impact of stress in self- and peer assessment. *Assessment and Evaluation in Higher Education. 30*(1), 51-63.

Popham, W. J. (2004). Why assessment illiteracy is professional suicide, *Educational Leadership,* 82-83.

Popham, W. J. (2002). *Classroom assessment: What teachers need to know* (3rd ed.). Boston: Allyn & Bacon.

Popham, W. J. (2001). *The truth about testing: An educator's call to action.* Alexandria, VA: Association for Supervision and Curriculum Development.

Porter, K. & Snipes, J. (2006). *The challenge of supporting change: Elementary student achievement and the Bay Area School Reform Collaborative's focal strategy (*ED494666). New York: Manpower Demonstration Research Corp.

Quinn, T. (2002). What principals should know and be able to do: Redefining leadership in the standards era. *Principal, 82*(1) 16-20.

Resnick, D. (1980). Minimum competency testing historically considered. *Review of Research in Education, 8*, 3-29.

Resnick, L. & Hall, M. (1998). Learning organizations for sustainable education reform. *Journal of the American Academy of Arts and Sciences, 127*(4), 89-118.

Resnick, L. & Klopfer, L. (Eds.). (1989). Toward the thinking curriculum: An overview. In L.B. Resnick & L.E. Klopfer (Eds.), *Toward the thinking curriculum: Current cognitive research.* Alexandria, VA: Association for Supervision and Curriculum Development.

Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education, 17*(1), 1-24.

Ross, J.A., Rolheiser, C. & Hogaboam-Gray, A. (2002). Influences on student cognition about evaluation. *Assessment in Education, 9*(1), 81-95.

Ruff, W. & Shoho, A. (2005). Understanding instructional leadership through the mental models of three elementary school principals. *Educational Administrative Quarterly, 41*(3), 554-577.

Rutherford, F. (1997). *Reflecting on Sputnik: Linking the past, present, and future of educational reform.* Retrieved October 13, 2006 from www.nas.edu/Sputnik/ruther1.htm

Sadler, D.R. (1998). Formative assessment: Revisiting the territory, *Assessment in Education: Principles, Policy & Practice,* 5(1), 1-7.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Schafer, W. (1993). Assessment literacy for teachers. *Theory into Practice, 32*(2), 118-126.

Schunk, (1996). Goal and self-evaluative influences during children's cognitive skill learning,

American Educational Research Journal, 33(2), 359-382.

Scriven, M. (1967). The methodology of evaluation. In R. F. Stake (Ed.), *American Educational Research Association monograph series: No. 1. Curriculum evaluation.* Chicago; Rand McNally.

Senk, S., Beckmann, C., & Thompson, D. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education, 28*(2), 187-215.

Sergiovanni, T. (1991) *The principalship: A reflective practice perspective* (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Shellard, E. (2005). How assessment data can improve instruction. *Principal, 84*(3), 30-32.

Shepard, L. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, 66-70.

Shepard, L. (2000a). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Shepard, L. (2000b). The role of classroom assessment in teaching and learning. (CSE Technical Report, 517). Los Angeles: Center for the Study of Evaluation

Shepard, L., Flexer, R., Hiebert, E., Marion, S., Mayfield, V., & Weston, T. (1996). Effects of introducing classroom performance assessments on student learning. *Educational Measurement: Issues and Practice, 15*, 7-18.

Sheppard, B. (1996). Exploring the transformational nature of instructional leadership. *The Alberta Journal of Educational Research, 42*(4), 325-344.

Sirin, S. R. (2005). Socioeconomic status and student achievement. A meta-analytic review of the research. *Review of Educational Research, 75*(3), 417-453.

Sloane, F. & Kelly, A. (2003). Issues in high-stakes testing programs. *Theory into Practice, 42*(1), 12-17.

Smith, W. & Andrews, R. (1989). *Instructional leadership: How principals make a difference*. Alexandria, VA: Association for Supervision and Curriculum Development.

Soles, D. (2001). *Sharing scoring guides.* Paper presented at the Annual Meeting of the Conference on College Composition and Communication in Denver, Colorado.

Solmon, L. & Podgursky, M. (2000). *The pros and cons of performance-based compensation* (Opinion Paper No. 120). Santa Monica, CA: Milken Family Foundation.

Starch, D. & Elliot, E. C. (1913). The reliability of grading high-school work in mathematics. *School Review, 21*, 254-259.

Stiggins, R. J. (2005). Assessment for learning: Building a culture of confident learners. In DuFour, R., Eaker, R., & DuFour, R. (Eds.), *On common ground: The power of professional learning communities (pp. 65-83).* Bloomington, Indiana: National Educational Service.

Stiggins, R. J., et al. (2005, September). *Assessment for learning defined*. Symposium conducted at the meeting of the ETS/Assessment Training Institute's International Conference: Promoting Sound Assessment in Every Classroom, Portland, Oregon.

Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan, 86(*1), 22-27.

Stiggins, R. J. (2002). *Assessment crisis: The absence of assessment FOR learning*. Retrieved February 21, 2006, from http://www.pdkintl.org/kappan/k0206sti.htm

Stiggins, R. J. (2001). *Student involved classroom assessment* (3rd ed.). Upper Saddle River, NJ:

Merrill Prentice-Hall.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan.*

Stiggins, R. J. (1997). *Classroom assessment.* Upper Saddle River, NJ: Merrill Prentice-Hall.

Stiggins, R. J. & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), p. 271-286.

Stiggins, R. J. & Conklin, N. (1992). *Investigating the practices of classroom assessment.* New York: State University of New York Press, Albany.

Stiggins, R. J., Frisbie, R. J., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice 8*(2), 5-14.

Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement, 26*, 233-246.

Stiggins, R. J., Rubel, E., & Quellmalz, E. (1986). *Measuring thinking skills in the classroom.* Washington, DC: NEA Professional Library.

System for Adult Basic Education Support (SABES) (2006). *Glossary of useful terms.* Retrieved April 11, 2007 from http://www.sabes.org/assessment/glossary.htm

Thattai, D. (2001). *A history of public education in the United States*. Retrieved October 17, 2006 from http://www.servintfree.net/~aidmn-ejournal/publications/2001-11/PublicEducationInTheUnitedStates

Valentine, J. & Bowman, M. (1988). Audit of principal effectiveness: A method for self-improvement. *NASSP Bulletin, 72*(508), 18-26.

Valentine, J., Clark, D., Hackmann, D. & Petzko, V. (2004). *Leadership for highly successful middle level schools, Volume II: A national study of leadership in middle level schools.* Reston, VA: National Association of Secondary School Principals.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press

Whiting, B., Van Burch, J. & Render, G. (1995). *Mastery learning in the classroom.* Paper presented at the Annual Meeting of the AERA. San Franscisco, 1995 available from ERIC ED 382688.

Wiggins, G. (2002). *Grant Wiggins on assessment.* Edutopia: The George Lucas Foundation. Retrieved December 17, 2006 from http://www.edutopia.org/php/interview.php?id=Art_935&key=005#paragraph3

Wiliam, D. (2000). The meanings and consequences of educational assessments. *Critical Quarterly, 42*(1), 105-127.

Wiliam, D. & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions as assessment? *British Educational Research Journal, 22*(5), 1-11.

Wiliam, D., Lee, C., Harrison, C. & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1), 49-65.

Wolf, M., Crosson, A., & Resnick, L. (2006). *Accountable talk in reading comprehension instruction.* CSE Technical Report 670, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and student testing.