

## CHAPTER II: REVIEW OF LITERATURE

### Assessment and Reform

Many groups have been commissioned to study reform and have often times recommended major changes in the existing educational structures. It is the purpose of much of our mandated assessment programs to produce educational change and to “improve the performance of pupils, teachers, and the educational system in general” (Airasian, 1993, p. 23). The “Back to the Basics” movement has also helped spur external testing programs. The problem has become actually identifying what the basics are. Some educators have suggested that children are the basics (Farr & Carey, 1986). Wraga (1994) confirms that educators are often assigned a peripheral role when politicians and business interests are busy changing schools. Politicians, in particular, work within a limited time frame and must make decisions limited to a period of relatively short duration. They push for changes and answers quickly, particularly if elections or other imminent political changes are near (Levin, 1991). Finn and Reparber (1992) relate that “taxpayers are no longer content to hand over their wallets and ask no questions” (p. 184). Policy makers are anxious to reform education and provide data to the voters that their reform initiatives are getting the job done (Farr & Carey, 1986). Shavelson et al (1992) caution that accountability systems which quickly follow political rhetoric may be just as likely to drive education in unwanted directions as the desired direction.

Politicians and policy makers support achievement tests as a critical and necessary feature of the accountability movement. They

believe that by examining scores the public can know whether schools are fulfilling their mission -- educating children successfully (Paris, 1992). While it may be possible for certain educational tests to reflect the skills that are important for today's student to possess, the standardized tests usually chosen by policy makers are given more powers than they were intended to have. They are used as instruments to select, assess, prescribe, reward, punish, and guide, not serve merely as indicators of what has been taught and learned (National Education Association, 1991). Standardized tests have become a political tool and educators have little recourse but to conform (Wraga, 1994). State mandated tests eliminate local discretion and generally provide information to a wide audience which continues to chip away at local control (Corbett & Wilson, 1991). The problems of achievement testing in America are unrecognized and unacknowledged by many people. Changes have often been mandated by policy makers whose "specific expertise may not exceed the vantage point offered by their former position of student in the educational process" (Page & Page, 1988, p. 1). Both the public and professional educators must become aware that the goals we are striving to meet are not adequately reflected in standardized test scores (Livingston, Castle, & Nations, 1989). They threaten the validity of the tests and "more importantly, threaten the motivation and learning of students and the effectiveness of teachers" (Paris, 1992, p. 96). Costa (1989) encourages us to redesign assessment from product-oriented assessments toward those designed to measure a process-oriented education. Corbett and Wilson (1989) believe that the pressure for uniformity and quick success "contradicts everything that is known about the process of improving

schools” (p. 25). Schools improve when all factors of the school, demographics, culture and organization, are considered and adequate time and technical assistance is offered. Our education must change to fit the demands of the information age and our conception of effective schools must broaden so that we can actively seek and promote alternative forms of assessment and accountability (Livingston et al, 1989).

### High-stakes Assessment

Tests which are used for important decisions are high-stakes assessments. These purposes may include promotion, graduation, and certification. The consequence does not need to be as serious as losing accreditation to be considered a high-stakes assessment. Publishing scores in the newspaper is considered to be a severe consequence by some teachers, schools and districts (Monsaas & Engelhard, 1994). Low-stakes tests generally impose no sanctions or rewards. Because of the importance attached to high-stakes tests their original purposes and limitations become obscured. They take on greater importance and significance than they merit leading to a narrowing of instruction and emphasizing the curriculum content that will be tested (Meisels, 1989). Madaus (1988) reports that research findings prior to 1988 indicated a minor role for traditional standardized test information but he reminds his reader that this was during the pre-‘reform with a test’ era. When tests are used to drive reform, many unintended behaviors and outcomes are likely to occur.

### Impact on Curriculum

State-mandated standardized tests have a profound influence on

the curriculum. As textbooks become geared to the test objectives, as time is allocated to the objectives covered by the test and objectives not tested are not taught, as the cost of testing paraphernalia takes away monies which could otherwise be spent on curriculum appropriate to the local school, the test actually becomes the foundation of the curriculum and the testing agency has become the controlling agent of the curriculum (Livingston et al, 1989; Meisels, 1989; Shepard, 1989). Farr and Carey (1986) expressed the fear that as tests become more influential in curriculum choice, schools will be little more “than test preparation academies where the curriculum becomes what can be tested” (p. 210).

#### Impact on Teachers and Instruction

When a district adopts a test structured around a particular instructional method, teachers often feel constrained to use that approach. For example, if a test for reading is a cloze test, one that requires the student to fill in blanks, the teacher may emphasize that technique in teaching. If the test requires a written response to an open ended question, the teacher will concentrate on written responses (Meisels, 1989). As teachers use a narrowed curriculum and teach with methods that may not be ethical or developmentally sound, a deprofessionalization of teaching occurs. Teachers find themselves in an awkward position when teacher judgment and non-test documentation of student learning can be invalidated by a single test score (Corbett & Wilson, 1989; Livingston et al, 1989).

#### Impact on Students

Livingston et al (1989) reported that “only rote learning takes place when teachers ‘pound’ information and skills into students who lack

readiness. Testing students on skills for which they lack readiness causes failure. Repeated failure breeds low self-esteem and negative attitudes toward learning, perhaps causing some students to drop out. Students become bored with repeated testing; some become cavalier about taking tests, increasing their likelihood of school failure” (p. 24). Madaus (1988) is concerned that school programs which use test results for grade-to-grade promotion will not improve student learning, but will eventually increase the dropout rate.

### Ethos of the School

The Nation at Risk describes the American promise:

All, regardless of race or class or economic status, are entitled to a fair chance and to the tools for developing their individual powers of mind and spirit to the utmost. This promise means that all children by virtue of their own efforts, competently guided, can hope to attain the mature and informed judgment needed to secure gainful employment, and to manage their own lives, thereby serving not only their own interest but also the progress of society itself (p. 8).

One can hear the spirit of the individual in this pledge. However, as standardization of education becomes the goal, a quality of sameness settles over all students. The same standardization of instruction begins to permeate teaching. Traditionally, teachers have a great deal of flexibility in deciding how to teach, what to teach, and when to teach it. Testing programs challenge this ethos. Tests dictate curriculum,

instructional techniques, even the sequence and order objectives are taught. Corbett & Wilson (1991) remind us that impact is not the same as improvement.

#### Impact on Businesses, Taxpayers, Parents.

Test scores have a general appeal to the public and enjoy a great deal of public acceptance. Tests have played a part in the lives of most citizens and because of that experience, are meaningful to individuals. They represent order and control. The language associated with testing is powerful. The connotations of words like standard, competent, excellence, objective, reliable, and valid symbolize important educational concepts. Tests are symbolic of values that society holds dear -- hard work and reward for effort (Airasian, 1993). Test results are generally reported at a local school board meeting and relayed to the press which, in turn, prints scores in the local paper. Particular circumstances which impact scores, such as socioeconomic differences and other school differences, get lost in the rhetoric that follows publication of scores (Farr & Carey, 1986). Parents, often ill-equipped to understand the scores reported for the schools in their neighborhoods, make use of these same results and request placement of their children in high scoring buildings. Real estate agents use test scores to advertise housing in certain districts, allowing the housing market, and sales commissions to adjust upward according to the scores. Commercial publishers sell achievement tests, scoring services, and data reports in order to make a profit. In 1993 it was estimated that state and local tests cost over \$1 billion per year (Stiggins, 1993). Added to the cost should be expense of commercial test training materials many districts employ to raise scores. The consequences of

testing policies are far more reaching than the designers of tests ever anticipated (Corbett & Wilson, 1989; Paris et al, 1991).

### Kansas State Assessments

When the State of Kansas adopted the Quality Performance Accreditation model for school reform the state also initiated a state-wide system of assessments designed to measure student improvement in selected areas and to drive instruction. In 1994 the State Board of Education issued the criteria which would be used to determine the accreditation status of each school at the conclusion of that school's Quality Performance Accreditation cycle.

One basis for accreditation is continuous improvement of performance by all student groups in academic areas over time. Performance by student groups on state assessments are measured against standards of excellence determined by the state as well as used to identify trends in achievement in individual buildings. Each school must also maintain or improve other indicators such as attendance rates, acts of violence and mastery of algebraic concepts.

Large scale assessments have traditionally been objective pen-and-pencil multiple choice tests (Stiggins, 1993). As we continue through the 90s, the Kansas State Department of Education has included standardized performance tests as well as more traditional types of tests. The Kansas Writing Assessment requires students to write an essay over a given period of time incorporating the process model of writing. These essays are scored by trained teachers using a six-trait analytical model rubric. The Kansas Reading Assessment features both narrative and expository text. The questions include multiple answer items and written

responses to more open-ended questions. Trained teachers score the written portions using a five-point rubric while the more objective portions of the test are sent to a testing center to be scored.

In addition to using the state assessments, schools were advised to consider a variety of additional measures to show continuous improvement toward identified goals. No one indicator will be used to determine accreditation status and schools are encouraged to consider multiple indices of student growth. Accreditation information furnished each school states “multiple measures collected over multiple years will be reviewed holistically in order to determine whether or not improvement has occurred. From this review, an accreditation recommendation will be made” (Kansas State Board of Education, 1993). Buildings have been encouraged to include both statistical data as well as qualitative information. Examples of statistical measurements include standardized norm-referenced achievement tests, state assessments, and district wide tests. Qualitative information might be teacher observations, performance ratings, portfolio assessments and projects. Because the state assessments are considered in the accreditation process, they are viewed as mandated measures. Other assessments are determined by each building. Measurements chosen by each school must be approved by the On-site team and the Kansas State Department of Education representative for that particular school. Because of the nature of the accreditation process, these measures, though self-determined in part, can be considered high-stakes measurements.

## School Improvement Plan Assessments

Like other schools and districts across the country, Kansas schools have many decisions to make when selecting measurements or assessments to show improvement in targeted areas. They may choose direct measures, indirect measures or a combination of the two.

A direct measure is one in which the student actually does what the assessment is presumed to measure. Writing an essay would be an example of a direct measure or assessment. Multiple choice questions about writing, such as those found in many achievement tests, constitute an indirect measure. It 'indirectly' measures the skill it is designed to assess. Stiggins (1982) notices advantages for both types. Direct assessments give information about an examinee's actual proficiency, stimulus and response are generally in agreement, exercises can be adapted to authentic circumstances, high face validity, and low test development costs. Indirect assessments offer high score reliability, relatively low test scoring costs, and a high degree of control over the nature of the skills tested. Each type also has disadvantages. Direct assessments may incur high scoring costs and there is a potential lack of uniformity of proficiencies assessed among examinees. Indirect assessments often lack authenticity, rely heavily on non-tested skills, and often lack face validity.

Most objective tests are indirect measures. An objective test item has one right answer known to the test-maker. Choosing one appropriate answer is only one indication of comprehension. There are many other ways to demonstrate proficiency. Students who are only 'allowed' to answer multiple choice questions are denied other ways to demonstrate

their abilities (Glazer, Searfoss & Gentile, 1988). Subjective measures involve judgment on the part of the evaluator. They often involve a range of answers or responses to a problem. Madaus (Brandt, 1989) believes this way of measuring to be excellent but requires a greater degree of trust in teachers than is expected with more objective measures.

As school improvement committees began the work of developing a school improvement plan for communications they looked for assessments their faculty could use to show improvement in the targeted area. There are many options available. The remainder of this chapter is devoted to a closer examination of the assessments chosen by Franklin County schools to show improvement in communication skills, generally reading and writing.

### Formal Measures

The most frequent type of tests are formal measures of reading. Formal tests have been given to a large number of students to establish a norm group. Individual and small group performances can then be compared to the performance of the norm group.

Standardized norm-referenced tests. Most buildings chose one of many nationally published standardized norm-referenced achievement tests as one of the indicators of growth. These tests generally cover the usual school subjects with reading and language arts being subtests of the total battery. According to Salmon-Cox (1981), teachers use achievement test information in several ways: as confirmation of what they already believe of students, to reflect on and guide instruction, to group and track students. Teachers are thought to use test results; yet, when teachers are asked about how they assess their students, they

seldom mention standardized tests.

Teacher attitudes towards traditional multiple-choice, standardized, norm-referenced tests are fairly well documented (Corbett & Wilson, 1991, Monsaas & Engelhard, 1994, Stiggins & Bridgeford, 1984). Stiggins and Bridgeford (1984) reported that teachers viewed traditional tests as “time-consuming, not matching their instruction, failing to reflect true student characteristics, and generally not meeting important instructional needs such as identifying material to teach or reteach” (p. 278). Corbett and Wilson (1991) concluded that statewide testing programs, generally based on a standardized test, force districts to act on results (a high-stakes test) but these actions do not represent what educators call ‘improvement’. Herman and Golan (cited in Nevo, 1995), found teachers did not believe that standardized testing was helping schools improve nor did it provide useful feedback to teachers or students.

Monsaas and Engelhard’s (1994) research involved comparing teachers’ attitudes toward standardized test preparation and actual behaviors. He found positive attitudes were negatively correlated with behaviors -- as pressure to improve test scores increased, so did the incidence of test-prep behaviors. George Madaus, in an interview with Brandt (1989), reports that “any move to school-based management will have to examine closely the implications of high-stakes measurement-driven instruction for school-based decision making” (p.27).

Young children place their trust in their teacher and assume testing is proper and appropriate, but older students report decreased motivation and feel less prepared to take the tests (Paris, Lawton, Turner, & Roth, 1991). Paris et al (1991) report that the results for low achievers become

increasing less valid. The “ironic consequence is that the scores become high-stakes confirmation of their lack of success” (p. 15). They also reported that students suffer from test anxiety, are cynical about tests, and may feel threatened, worried, or bored by standardized tests.

#### Criterion-referenced tests.

Criterion-referenced reading tests are generally considered very specific regarding the skills tested, the number of items developed and the relation of test objectives with those of the curriculum. Scores are interpreted as a measure of performance compared to a standard. Using a criterion-referenced test, a student’s performance is reported in terms of content that has been mastered. A student is compared to a standard of proficiency rather than other students who took the test. Criterion-referenced tests focus on what a student knows and can do (Farr & Carey, 1986). Anastasi (1988) notes that criterion-referenced testing is best suited for testing basic skills at the elementary level. Mastery testing is not appropriate for such skills as critical thinking or creativity where students’ achievement may be almost limitless. Criterion-referenced tests assume a linear model of learning which means knowledge is built by placing one discrete fact atop another, much in the way a brick wall is built. One cannot work at the top until the bottom bricks have been laid (Shepard, 1991).

#### Publishers’ textbook tests.

Another formal measure of reading is the text-related test. Farr and Carey (1986) describe basal reader tests as an assessment that falls between formal standardized tests and informal observations made by the teacher. They are typically administered to pupils periodically and results

are often reported to the principal, filed in cumulative records and reported to parents. These end-of-unit tests generally measure the students' acquisition of skills during a particular instructional event--the completion of the study of a book, chapter or unit. Stiggins and Bridgeford (1985) explained that nearly half of the teachers studied report using published tests with ease with the remainder not using them at all. Stansfield (1976) noted that two-thirds of the foreign language teachers he studied use only tests they themselves have prepared. Foreign language department chairpersons surveyed rejected textbook tests overwhelmingly. Yeh, Herman, and Rudner (1981) noted that teachers with classroom aides reported greater use of curriculum-embedded tests than teachers working without aide assistance. Farr and Carey (1986) believe that these tests may be the "driving force behind classroom reading instruction" (p. 204), but very little evidence regarding use and technical aspects is available.

Teachers who reject published tests, including both standardized tests and text-related tests, express concerns about student reactions. They tended to view the tests as invalid, undependable, too long, etc. and believed that the tests were generally not helpful to students. In the Stiggins and Bridgeford (1985) study, teachers' reactions to their own paper and pencil objective tests, published tests, and performance assessments were compared. Published tests generated the most negative comments and were viewed by teachers as interfering with instruction.

### Informal Measures

Formal measures of reading almost always need to be

supplemented with informal measures. Informal measures usually contain more error than formal measures but they supply useful information for teaching applications. They generally involve the actual materials and tasks used in the classroom and the teacher is able to generate a more realistic view of how the student performs in the educational setting. Paris (1992) expresses strong opinions about educational assessment. It must “be changed to serve students, teachers, and parents rather than political purposes. The public’s mistrust of teachers’ professional judgment must be allayed by better assessments. Teachers’ frustration and hopelessness in the face of externally imposed high-stakes testing must be abolished. Students’ developing discontent and disillusionment with educational assessment must be rectified” (p. 104). Paris et al (1991) also believe that performance and portfolio assessment will help remedy many of the shortcomings of standardized tests.

Performance assessments share many important characteristics. Students “are called upon to apply the skills and knowledge they have learned. Second, performance assessment involves completion of a specified task (or tasks) in the context of real or simulated assessment exercises. Third, the assessment task or product completed by the examinee is observed and rated with respect to specified criteria in accordance with specified procedures” (Stiggins & Bridgeford, 1985, p. 273). In general, teachers’ concerns about performance assessments focus on accuracy, defining levels of performance, the need to be objective, providing diagnostic information, measuring growth, and the time demands of conducting performance assessment.

Many states are including performance-based assessments and portfolios in their state testing programs (Aschbacher, 1994). Kansas state assessments range from single answer multiple choice (portions of the Math Assessment) to cooperative group work scored with a rubric (fifth grade Science Assessment) to a direct measure of writing, also scored with a rubric.

#### Portfolio assessment.

Several schools in Franklin County have chosen portfolio assessment or a collection of work samples to document student learning. Portfolio assessment exists when the purpose is defined, guidelines for the inclusion of work in a portfolio have been determined and criteria for either the individual pieces or the collection as a whole have been identified. Portfolios, by their very nature, lend themselves to large variations in implementation, as well as in scoring (Benoit & Yang, 1996). Koretz, Stecher, Klein, and McCaffrey (1994) describe difficulties with reliability and validity in portfolio large-scale assessments. They advocate greater standardization of tasks, revision processes, and preparation but acknowledge that such standardization runs contrary to many of the goals of portfolio assessments. Teachers report that student interest and learning increase as they assume ownership of their own work. Parents also express interest in the rubrics associated with portfolio collections (Cortez, 1994). Teachers must commit time and staff development to portfolio development activities if the assessment is to be of value in improving classroom instruction (Benoit & Yang, 1996). While the costs involved include time, increased stress on teachers, and financial investment, Vermont teachers generally believed the portfolio program

was a worthwhile burden (Koretz et al, 1994).

### Accelerated Reader™.

Accelerated Reader™ is a computer software package which provides students computer-generated multiple-choice tests to assess their comprehension of books they have read. Each student generally self-administers a test on one of over 1000 titles available (Turner, 1993). Teacher researchers reported that the primary benefit of Accelerated Reader™ is as an extrinsic motivator for encouraging students to read. Teachers were generally positive toward Accelerated Reader™ (McKnight, 1992; Peak & Dewalt, 1993; Turner, 1993).

### Postscript

“Do reading tests really measure reading?” Farr and Carey (1986) suggest that when we consider the reading that human beings do for their own reasons the answer must be “no.” “Despite what one often reads in research articles, reading is not what reading tests test. If test consumers understood, accepted and acted on the basis of this basic fact, we might witness significant improvements in the use of reading tests” (p. 16). Nickerson (1989) summarized attitudes toward assessment held by teachers. “Some people object to testing in principle, some believe too little is done, others feel testing can serve important purposes but that it is often used in inappropriate or objectionable ways” (p. 6). Frederiksen (1994) looks toward a future when tests can be used during the instructional process. Aschbacher (1994) recommends that a much better appreciation of educators’ understandings of and responses to new assessments is necessary if new forms of assessment and instruction are to improve our schools.