

# Designing a Sustainable Standards-based Assessment System

by Don Burger

© Mid-continent Regional Educational Laboratory, Inc. 1998

---

## Introduction

The national education goals adopted by President Bush and the nation's governors at the Charlottesville, Virginia "Education Summit" in 1990 gave rise to the development of national standards in many content areas. The National Council of Teachers of Mathematics (NCTM) had taken the lead by issuing its "Curriculum and Evaluation Standards for School Mathematics" in 1989. Other domains such as science created committees to attempt to resolve divergent views within content domains and find ways to reach consensus about what students should know and be able to do. Educators in history, social studies, civics, geography, language arts, the arts and physical education have been exploring areas of common ground and areas of divergent opinion in an effort to reach consensus on standards.

The recent re-authorization of the Elementary and Secondary Education Act--now the Improving America's Schools Act (IASA)--increased the importance of having standards for each school receiving federal funds. While some K-12 schools and districts are willing to accept the challenge of systemic change, others will develop goals or content standards and insert them in front of their existing curriculum guides without making any changes in the educational system. The same schools will then deem themselves in the forefront of the standards movement and aligned with national standards.

However, designing and implementing a sustainable standards-based system that consistently yields high student achievement involves more than setting and measuring academic goals. The process of changing to a "standards" frame of reference goes much deeper. Changing to a standards-based system provides an opportunity to re-examine the organizational elements (Cordell & Waters, 1993) of a school system: fundamental purpose, principles, policies, processes, practices, programs and procedures. Standards present an opportunity to examine or clarify these organizational elements as they are viewed by all the groups within the school community--teachers, school administrators, other district staff members, parents and other community members. Establishing clarity on the organizational elements--e.g., purpose, principles, and policies--may do more to help schools and their communities re-establish trust in public education than many of the current reform efforts which tinker with how schools work--or don't work.

Figure 1 is an attempt to illustrate the interrelationships of the organizational elements with classrooms, schools, districts, and communities as systems. The top of the diagram represents these four systems. The boundaries between these systems are permeable. Interaction between systems occurs most frequently with the systems closest to one another. Within each system, there are a number of organizational elements that direct and validate the actions taken by the system and its members. These organizational elements--purpose, principles, policies, processes, practices, programs, and procedures--are represented in Figure 1 by the diagonal slice (Waters, personal communications, 1995). Each of the organizational elements interacts with the others, with those that are more proximal interacting most. The outer layers of procedures, programs, practices, and processes are most permeable, are the easiest to change, and

offer the least resistance to change. Consultants find changes aimed at the procedural and programmatic levels are the easiest to implement. Organizational elements toward the center of the diagram, e.g., purpose and principles, are less permeable, are much more difficult to change, and involve much greater resistance to the changes.

While educational reforms that tinker with the outer layers are easiest to accomplish, they have the least meaningful, sustained impact on student learning. The literature is replete with efforts aimed at practices, programs or procedures that have little sustained significant impact on student learning. To achieve significant and sustainable improvements in student learning requires changes in the fundamental core elements of the system. This article will describe each of the organizational elements and will address the ways in which each of these elements influence the design of a sustained standards-based assessment system.

Each organizational element has a role in directing how schools are run, how students are taught, and how students are tested. For example, the impact of the traditional purpose of schooling was to sort and select students. Traditional purpose was supported by the belief that not all students can learn, that education should "tease out" the very best, and that some students should fail. Consequently, principles, policies, processes, practices, programs and procedures were developed in which all students were taught the same way, given the same amount of time, and tested with assessments based on the "normal" (bell-shaped) curve.

In contrast to traditional beliefs, researchers are finding that, although students do learn at different rates and in different ways, virtually all students can learn. Given appropriate time and instruction and clear expectations, many of the students previously written off can meet or exceed rigorous academic standards. Classrooms, schools and districts that have been successful in creating educational systems in which all students are learning have changed more than educational goals and tests. Successful sites have changed the core of the system: purpose.

---

## **Purpose**

Purpose is at the core of each system and is most difficult to change. School staff frequently reflect many diverse views about the purpose of schooling. The combinations and interactions of the views of stakeholders form the culture of the school and underlie principles; direct policy; and validate processes, practices, programs and procedures.

Changing purpose is a slow and painful process. Members of the school/community become uncomfortable when different perspectives challenge their own. Resistance to change is always present. Resistance is the system's way of remaining the same. Staff who are very comfortable with things just the way they are, are comfortable because their beliefs match those of the system. For example, staff who believe that some students cannot learn are very comfortable moving forward students who are failing.

The central purpose for schooling is at the core of the system. Purpose is reflected by the question, "Who is expected to learn?" Traditionally, schools were designed to serve the best and the brightest, sorting out the others for manual labor jobs. Traditionally, white children of high socioeconomic status (SES) were expected to do best. The traditional educational system was designed to further that end. Standards-based

education introduces a different purpose, one that is focused specifically on sustained student achievement for all students--regardless of gender, race/ethnicity, or socioeconomic status. Table 1 contrasts traditional and standards-based education systems in terms of purposes of schooling and impacts of their contrasting purposes.

**Table 1**

<b>PURPOSE</b>	<b>FROM</b>	<b>TO</b>
<b>WHO IS EXPECTED TO LEARN?</b>	Learning is the birthright of white high SES students.	All students can learn given appropriate time and instruction.
<b>IMPACT</b>	<p>Not all students can learn. The best and brightest survive.</p> <p>Equity means students have equal access to programs that result in unequal performance by groups.</p>	<p>Each student is expected to meet or exceed standards.</p> <p>Equity means that each student receives the instruction and time required to reach the standard. While all students reach the standard, differences in performance still exist.</p>

## Principles

Principles must be congruent with the "core purpose" of the system. Principles provide the direction and guidance for the system much as a compass always registers north regardless of one's position (Covey, 1991). Guidance is precisely the value of principles. In assessment, there will always be new content standards and new ways of measuring those standards. It is the principles that will provide the guidelines and parameters for selecting the new or replacement assessment systems.

Seven design principles are required to create an assessment system that produces significant and sustainable improvement in student learning. These design principles are more than a map or guide; they provide clear direction. The seven design principles build an assessment system that is accountable, flexible, standards based, anchored, ongoing, disaggregated, and transition focused (Cordell & Waters, 1993). When tests and assessments change, it is the design principles that remain constant. Any new assessment that is incongruent with the design principles will be purged by the system. Table 2 below shows the differences between traditional and standards-based systems on all of these design principles.

**Table 2**

<b>PRINCIPLE</b>	<b>FROM</b>	<b>TO</b>
<b>1. ACCOUNTABLE</b>	Test data are not used for instruction or verification of student learning.	Commitment that all students will meet standards by the time they make their transition from level to level or graduate.

<b>2. FLEXIBLE</b>	Testing is a fixed, grade-leveled event that accommodates variation in student learning by expecting a bell-shaped distribution of scores.	Testing is flexible. Students challenge the test when they have demonstrated they are ready to be successful. Differential learning rates are accommodated by offering the tests over groups of grades rather than grade levels.
<b>3. STANDARDS-BASED</b>	Student performance is compared to that of other students even though some students may not have received instruction on the material tested.	Student learning and performance are measured against a standard through the use of valid and reliable instruments.
<b>4. ANCHORED</b>	Student performance is compared to the performance of other students on an average national curriculum.	Internal district standards are tied to acceptable external standards through the correlation of performance on the performance assessment to performance on a traditional norm-referenced achievement test.
<b>5. ONGOING</b>	Testing is an annual, one-time event, usually during October or April.	Testing is a continual process that provides student performance data to teachers and students in "real time."
<b>6. DISAGGREGATED</b>	School and district scores are reported as single mean percentiles. Scores are not reported by gender, race/ethnicity or socioeconomic group.	Scores are disaggregated by gender, race/ethnicity or socioeconomic status and are publicly reported. However, schools are not compared.
<b>7. TRANSITION-FOCUSED</b>	Since a normal distribution of scores is expected, there are no expectations that all students will be prepared to be successful at the next level. Students are socially promoted.	Clear expectations are defined for students to accomplish prior to moving from one level of the organization to another and graduating.

Systems that design for *accountability* commit to the concept that all students will meet standards. Traditionally, education has sorted and selected students (core purpose), much like separating cream from milk. The "cream" were destined for higher education and the remainder for the work force required in the industrial age. The traditional system filled the need for a large common labor force; consequently, students were identified through a failure process.

The concept of educating all students will be rejected by systems whose core purpose is sorting and selecting. Schools or districts must change purpose before education for all can be adopted by the organization. Rather than test students at fixed grade levels,

standards-based systems are *flexible*. Students challenge the tests after they have received instruction and when they have demonstrated they are ready to be successful. Testing becomes a success experience rather than a failure experience.

*Standards-based* systems bring clear focus on high but achievable targets for students to meet rather than comparing students to the performance of other students. Content standards precisely describe what students are expected to know and be able to do. When this information is shared with students, to no one's surprise, the students meet or exceed the standards.

The sustainable system is *anchored* against measures stakeholders have viewed as valid and reliable indicators. Standardized norm-referenced tests have become the de facto standard for American public education. Any replacement measurement system must be superior to the previously accepted system. As a beginning point, evidence about validity can be gathered by comparing the performance of the same group of students on both the norm-referenced test and the replacement measure. Standards-based or criterion-referenced assessments can be cross referenced to standardized norm-referenced tests by conducting a concurrent validity study (Burger & Burger, 1993). Students complete both a standardized norm-referenced test and a criterion- or standards-based test measuring a similar domain (reading for example). Statistical analysis will provide both correlational data and the relationship of performance standards to percentile ranks scores. The linkage between the two assessments can assure parents that the new assessments are rigorous and that the performance standards are worthy.

Sustainable standards-based systems change testing from an event in October or April to a continual and *ongoing* part of the instruction and assessment process. In the traditional educational system, the most efficient method of assessment is to test everyone at the same time. Standardized norm-referenced tests provide comparisons with other students in the same grades in the month of October, because teachers are still working on skills students have lost over the summer, or April, because students begin thinking about summer vacation in May. Assessment in standards-based systems is not limited by those parameters. Since the standard is fixed, students may attempt the test anytime they are ready to be successful. Assessment may not be an event at all. Assessment can be delivered as part of the instruction and assessment process.

Quality and equity, described by *disaggregated* student test data, are the basis for school improvement planning. Traditionally, quality was reserved for the top five or ten percent and equity meant that all students had an equal opportunity to participate. In other words, quality and equity focused on the "input" side of the system. Standards-based systems change the focus of quality and equity to the "output" side of the system. Quality is the evidence that all disaggregated data reflect attainment of high standards by students in all groups (race/ethnicity, gender, socioeconomic status). Equity is the evidence that there is no difference from one group to another in the percent of students meeting or exceeding standards. Disaggregated test data provide the evidence for quality and equity.

The sustainable standards-based system defines clear expectations for students to accomplish before transitioning, moving from one level of the system to another or graduating. The *transition focus* informs student, teachers, and parents about the status of the student on the knowledge and skills required to be successful at the next level.

Every staff member is responsible for seeing that all transitioning students have met or exceeded the standard rather than just the teachers in the grade level being measured as is common in traditional systems.

---

## **Policies**

Policies are the third element of the system. The most successful policies are those that match the purpose and principles of the system. Occasionally, policies that do not match the purposes or principles of the system are required by state or federal legislation. Those policies either cause chaos in the system or are enacted but never practiced. Careful examination of a district's policy manuals usually reveals many policies that are never practiced because they are incongruent with the purpose and principles of the system. Policies are required to maintain any assessment system. Assessment policies required in a sustainable assessment system include those that define and set standards and those that determine who, if anyone, is responsible/accountable for learning.

## **Standards:**

The term "standard" has been used synonymously to refer to curriculum standards, content standards and performance standards. Standards have come to mean many different things to many different people. Careful listening is required to determine which standards are being discussed. Kendall & Marzano (1995) distinguish curriculum standards, content standards, benchmarks and performance standards. Curriculum standards, they explain, "are best characterized as descriptions of what should take place in the classroom; as such, they address instructional techniques, recommended activities, and various modes of presentation" (p. 20). Content standards describe what students should know or be able to do. National groups developing standards--such as the National Council of Teachers of Mathematics (NCTM) and the National Committee on Science Education Standards--have mixed both curriculum and content standards in their standards frameworks. The term benchmark is used to describe the application of a content standard at a grade level or set of grade levels. Performance standards refers to the quality of the performance deemed acceptable for each content standard. Performance standards will be discussed in more detail along with rubrics in the section labeled "practice."

In addition, content standards have been viewed from two perspectives. The "literacy" model suggests that all students should meet or exceed performance standards in each content standard. "At the literacy end of the continuum, standards might be described as the minimum requirements of knowledge and skill students should know and be able to do to function well as adults of the 21st century" (Kendall & Marzano, 1995, pp. 13-14). The "expertise" model holds up the best examples to strive for, while expecting only the best to succeed. "At the 'expertise' end of the continuum, standards are described in terms of the knowledge and skills that, once acquired, would render students 'mini-experts' in every field" (Kendall & Marzano, 1995, p. 14). The standards-based system described in this article uses a literacy model approach in which all students would be expected to meet or exceed performance standards on each content standard.

The process of defining and establishing content standards in policy provides an opportunity to fundamentally question what we believe students should learn and to

reorganize the efforts of schooling to accomplish the goals. Table 3 contrasts how traditional and standards-based systems approach policy involving what students should learn. Schools and districts that adopt content standards that are incongruent with their purpose and principles will find that adopting content standards has made no difference in student learning, dropout rates or graduation rates. Traditional purpose and principles will resist changes to content standards at the policy level. Standards "too shall pass" just as have all the other initiatives which were very different from core purpose and principles.

**Table 3**

POLICY	FROM	TO
<b>WHAT SHOULD BE LEARNED?</b>	Curriculum adoption cycles, e.g., seven year cycles, determine when content area curriculum will be reviewed and new textbooks adopted.	Members of the school/community reach consensus on content standards, determine what is learned at each organizational level.
<b>IMPACT</b>	<p>A publisher's textbook series is adopted. The textbook becomes the de facto curriculum and teachers teach to the textbook.</p> <p>Fixed time - 180 class days and 45-50 minute class periods. All students proceed together regardless of learning.</p> <p>Teachers prepare one instructional delivery technique.</p>	<p>A variety of materials are used which provide experiences aligned with content standards. Teachers teach to content standards.</p> <p>Time varies depending upon student learning. Some students finish early and proceed to more challenging tasks. Some students receive more instructional time.</p> <p>Instruction is varied. Teachers prepare a variety of approaches that reach all students.</p>

**Stakes:**

In educational assessment, "stakes" refers to the consequences if learning does or does not take place. Who is accountable for learning? This is a question that must be resolved at the policy level of the organization. The easiest way to determine if the assessment system has stakes is to identify who is accountable for learning. The options are: no one (the most popular response); schools, but not teachers or students; or some combination of schools, teachers, and students. If there are no stakes for students, staff, schools, districts or states, i.e., no accountability, then the least expensive testing system without regard for validity, reliability and generalizability will suffice. However, if the system is accountable and stakes are used, then the issues of validity, reliability, and generalizability are extremely important.

"Stakes" can take different forms. Some districts and states have developed systems in which the schools, but not the staff or the students, are responsible for student learning. Students are asked to give effort on testing in which they have little or no interest and

which has no consequences. Is it any wonder we do not see superior performances? We are asking students to give their best effort on tests that we do not value enough to hold ourselves or the students accountable for the learning demonstrated.

"High stakes" systems are those in which promotion, certification or recognition can be denied based on performance or achievement as documented by assessment results. High stakes for students might require demonstrated competency before a diploma is issued. A medium stakes example ensues when diplomas are issued for class credits and "endorsements" are added for demonstrated competency. "High stakes" for staff would result if promotion, recognition or certification for staff were determined by rates of success in moving students to and beyond performance standards. "High stakes" might mean reassignment to another level where the person might be more effective or it might mean finding more suitable employment. "High stakes" for schools might require a change in the organizational structure for schools that are unable to move students to and beyond performance standards.

The assessment system must be properly prepared if "high stakes" are imposed for students or staff. The district must be able to prove that (a) what was tested was taught, (b) the reliability of the assessment exceeds  $r = .90$ , (c) students had more than one opportunity per year to attempt the assessment, (d) the tests are fair and free from bias, and (e) the knowledge and skills for which students were held accountable are really necessary for students to succeed (Herman, Aschbacher & Winters, 1992; Merhens & Popham, 1992; Phillips, 1993). Table 4 contrasts traditional and standards-based assessment systems with regard to stakes.

**Table 4**

<b>POLICY</b>	<b>FROM</b>	<b>TO</b>
<b>WHO IS ACCOUNTABLE FOR LEARNING?</b>	No one in the system is held accountable for individual student learning because it is the fault of students who do not want to learn or parents who do not make their student learn. School staff are not responsible.	The conditions for success lie within the school/community. School staff, students, parents and the school/community collaboratively are responsible for learning. School staff and students are directly responsible.
<b>IMPACT</b>	Since no one is held responsible for academic outcomes, mean or median test scores continue to be what they have been.	Students, parents and staff expect that all students will meet or exceed standards. All share responsibility for learning.

---

## **Processes**

In the context of organizations as systems, processes describe how purpose, principles, and policies will be delivered. Within the educational system, many processes must be defined to reach the organization's goals. The delivery of instruction is an essential part of the processes dimension. New theories about student learning and instruction that are

not congruent with the core values of the system are resisted. Table 5 contrasts the differences in processes between traditional and standards-based approaches.

**Table 5**

<b>PROCESSES</b>	<b>FROM</b>	<b>TO</b>
<b>HOW DO STUDENTS LEARN?</b>	Students learn in a linear fashion at the same rate.	Learning is uniquely individual to the student, non-linear, and based upon previous learning.
<b>IMPACT</b>	<p>Instruction is teacher centered.</p> <p>Learning is passive. Desks are placed in neat rows. Students work individually, orderly and quietly.</p> <p>Lecture is the primary mode of delivery.</p> <p>The distribution of student scores resembles the normal curve. Less than 10% do top quality work and failure is acceptable.</p>	<p>Instruction is tailored to the unique needs of the students.</p> <p>Learning is an active process. Arrangement of students from individual to group work changes continually throughout the day.</p> <p>Teachers are facilitators rather than dispensers of knowledge.</p> <p>All students meet or exceed performance standards.</p>

**District level accountability tests:**

A series of process level questions that must be answered involve the district accountability assessment system. What role will the accountability assessment play in the assessment system? What instructional decisions will be made based on results of the district level accountability test? These are "process" decisions that are directed by purpose, principles and policy. In turn, process decisions about the role of accountability assessment impact district assets, i.e., time, effort, and money.

Standards-based assessment systems that are used to inform instruction; require valid, reliable, generalizable tests; and provide immediate feedback to students and teachers cost more money. Typically, less than one percent of a district's budget is spent on assessment for accountability. Requests for increased funds must compete with other policy, process, practice, and program requests, such as wage and salary negotiations, new educational programs, changing curriculum and facilities. How stakeholders view the role of accountability assessment will determine how it is used and how it is financed. Table 6 contrasts the role of accountability assessment in traditional and standards-based systems.

**Table 6**

<b>PROCESSES</b>	<b>FROM</b>	<b>TO</b>
<b>WHAT IS THE ROLE OF ACCOUNTABILITY ASSESSMENT?</b>	Data are not used for instruction since the tests are not accurate measures of the taught curriculum.	Accountability assessment serves as a valid and reliable measure of the content standards.
<b>IMPACT</b>	<p>NRTs provide the best indicator of student learning given time and money.</p> <p>Tests and test scores are not valued by teachers or students.</p>	<p>Standards-based test data are an integral component in instruction. Test scores count and validate learning in the classroom.</p> <p>Teachers and students understand and value district accountability assessments.</p>

Traditionally, standardized norm-referenced tests (NRTs) have been used as the school and district accountability measure. While NRTs do an adequate job of comparing students to other students on basic skills, they do not measure student mastery of content standards. Depending on the definition of what students must know and/or be able to do established in content standards, NRTs may not be valid measures of the content standards. One factor contributing to the lack of validity of an NRT is the method of item selection. NRT items are selected based on their ability to make distinctions among students. Items missed or passed by most all students are not retained because they do not discriminate among students. Consequently, some content standards may not be measured. Emerging content standards offer schools, districts and states an opportunity to check the validity of the tests being used for accountability. Some parents, especially high socioeconomic status parents, value knowing that their student's performance compared well to the performance of other students. But what is the quality of the comparison? Since the test had no bearing on the students who participated in the norming sample, how much effort did those students give? What is the quality of performance at the fiftieth percentile or the ninety-sixth percentile? NRTs do not provide an answer. What does it mean when a student scores above XX percentile when compared to other students who did not care about the test? Schools and districts aiming for a score just above the mean may find the fiftieth percentile not a very worthy target.

If the accountability tests are to align with and measure content standards, criterion-referenced assessments (CRTs) that are valid, reliable and generalizable must be found or developed (Guskey, 1994). CRTs compare student performance to established criteria rather than to the performance of other students. CRTs allow all students who have acquired skills and knowledge to receive high scores. It is important to resolve the format of the district accountability test first, i.e., NRT or CRT, because non-alignment may still occur between district accountability assessments and classroom assessments.

## Test formats:

Test formats are another aspect of "processes." Table 7 contrasts processes involving test formats in traditional and standards-based assessment systems.

**Table 7**

<b>PROCESS</b>	<b>FROM</b>	<b>TO</b>
<b>WHAT WILL BE THE TEST FORMAT OF THE ACCOUNTABILITY TEST?</b>	Test formats are generally the same, the fastest to administer and the easiest to score.	Test formats must be appropriate measures of the content standards and be affordable for the school, district or state.
<b>IMPACT</b>	<p>The issue of validity is seldom raised or discussed. The tests are assumed to be valid.</p> <p>Establishing reliability on performance assessments is sacrificed for expediency.</p> <p>Single test formats are the rule.</p> <p>Standardization may or may not be used depending on the need to aggregate data.</p>	<p>There is a formal process for determining validity within the context of the content standards and budget.</p> <p>Formal reliability is a primary concern whether or not high stakes are used.</p> <p>A variety of performances will be used depending on the content standards and budgetary parameters.</p> <p>Standardization is essential because data are disaggregated to ensure comparability and to gauge quality and equity.</p>

Portfolios, performance assessment and authentic assessment are the current trend in student assessment. Machine scoreable, multiple choice formats, and tests that require the use of paper and pencil only have been criticized because they do not reflect practices in the "real world." However, there are "high stakes" multiple choice and paper and pencil formats that directly impact students and adults. The American College Test (ACT), Scholastic Achievement Test (SAT) and the Graduate Record Examination (GRE) are high stakes tests that use multiple choice and paper and pencil formats which college bound students face in the real world. Employers use paper and pencil tests as screening devices for employment which non-college bound students face in their real world. Paper and pencil and machine scoreable formats are just as authentic as other tasks performed in the real world.

The terms standardization and standardized norm-referenced test have come to mean the same thing, when in fact they are different. Standardization refers to the format and to administration procedures where all students in the comparison group take the same

test under the same test administration procedures. The advantage of standardization occurs when scores are aggregated or disaggregated or when issues regarding equating tests are raised. It is more difficult to compare students to students or students to standards when formats and administration procedures vary. Standards-based assessments also can be standardized by administering the assessment in the same way to all students (Hymes, Chafin, & Gonder, 1991). Likewise, performance assessments, portfolios and criterion-referenced assessments can all be standardized if process and procedures are the same. The first consideration in developing a test format is the content standard itself. For example, some content standards ask for pure recall of facts or knowledge. A knowledge-based, paper and pencil test is an excellent measure of content standards that ask for recall of facts. In contrast, content standards that focus on processes usually require a performance or demonstration. However, instances will occur when paper and pencil proxies might be used rather than performance assessments. These instances will occur when validity, reliability, generalizability or cost make the performance assessment prohibitive. In sum, a variety of assessment formats is desirable.

---

## **Practices**

In a sustainable standards-based assessment system, the practice element includes decisions about the number of performance levels, performance standards, and the role of classroom assessment.

### **Performance levels:**

The use of performance tasks as learning tools has become a popular method of integrating instruction and assessment. When a performance task is used, the quality of student performance is judged against a predetermined rating scale (rubric). Performance levels refer to the ranges of ratings of a performance task. Schools, districts and states use different performance rating schemes. Some states have two ratings: "pass" and "no pass." Others make finer distinctions about quality by using ratings that number from three to nine levels. The variety of rating levels reflects the many "practices" of schools and districts. Table 8 contrasts how traditional and standards-based systems approach performance levels.

**Table 8**

<b>PRACTICE</b>	<b>FROM</b>	<b>TO</b>
<b>HOW MANY PERFORMANCE LEVELS?</b>	The number of performance levels is viewed independently from reliability.	The number of performance levels must not compromise the reliability of the test.
<b>IMPACT</b>	<p>A number of performance levels are designated below the performance standard so that growth can be shown when the student did not achieve the performance standard.</p> <p>Test reliability is not an issue.</p>	<p>Usually one performance level exists below the performance standard. Students are rewarded when they meet or exceed expectations.</p> <p>Test reliability is a critical issue.</p>

Stakes also play a role in determining the number of performance levels. The percent of rater disagreements increases as the number of distinctions increases. Since high stakes systems require high inter-rater agreement, a lower number of performance levels is more effective. Four levels are a very common number of rankings. However, with any test, it makes little sense to report scores to students if the judgment is accurate only 50%, 60% or 75% of the time. High inter-rater agreement (reliability) is essential and is independent of the policy decision regarding stakes.

**Performance standards:**

Another part of the practice element is the process for setting the performance standards. While the performance standards may be set in district policy, the process of setting performance standards is a practice. A performance standard defines the quality of an acceptable performance. One performance level frequently is selected as an acceptable level of performance or a performance standard. While the performance standard is the least acceptable performance, it should not be viewed as a "minimum competency," the concept used in the 1970s. Schools and districts have developed different terminology that indicates whether student performance meets or exceeds expectations for the level tested. Performance standards should set high but achievable expectations for students. Performance levels should not be set so low that everyone meets or exceeds the performance standard. In standards-based systems, the school/community stakeholders are invited to the table to set district expectations for graduation and for moving from one level of the system to the next. Who sets the performance standards is an important decision at the "practice" level of the system. Table 9 shows how traditional and standards-based systems approach this practice issue.

**Table 9**

<b>PRACTICE</b>	<b>FROM</b>	<b>TO</b>
<b>WHO SETS PERFORMANCE STANDARDS?</b>	Each teacher sets the performance standards for his/her classroom.	Setting performance standards is a public process which involves representatives from the entire school/community.
<b>IMPACT</b>	Teachers set expectations for their classrooms individually.  Ratings and expectations vary from classroom to classroom. Class grades mean different things.	Groups of people representing the school/community examine performances and set expectations for all students.  Fixed clear targets are set for students to meet or exceed.

**Classroom assessment:**

Classroom assessment is another component of the practice element. Classroom assessments play a critical role in a standards-based system. Teachers need tools to make minute-by-minute instructional decisions for each student. Checklists, portfolios, teacher observations, and teacher made tests or tasks are the teachers' primary assessment tools. While classroom assessments may not use the exact same items or tasks as the district level accountability tests, they should be measuring the same knowledge or skills in approximately the same format as the district accountability test. Alignment of both classroom assessments and district level accountability assessments with content standards is essential. At the "practice" element level, classroom assessments are aligned with content standards and district level assessments for accountability.

Classroom assessment is most effective if what gets taught gets tested; if classroom assessment is aligned with district level accountability assessment and content standards; and if all these are congruent with purpose, principles, policy and practice. Aligned classroom assessment enables the teachers to make instructional decisions for students on a continual basis. Classroom assessments allow students to practice skills from simple to complex and to integrate those skills in meaningful ways. Students must know what skills they currently have and what they are expected to do in order to meet or exceed the standard. Since classroom and accountability assessment are aligned, there is no time wasted preparing for tests that occur only in October or April.

Teachers enjoy more latitude in the formats classroom assessment can take when district level accountability assessments are legally defensible. Short and long term individual and group performances, projects and portfolios are better suited to classroom assessment than to district level accountability assessment. More time can be devoted to assessments that take longer than a class period as assessment becomes part of the instructional process. Student self-evaluation can play a substantial role in classroom assessment. However, students will be more successful on district level accountability

assessments if the classroom assessments are similar in format (Herman, Aschbacher & Winters, 1992).

Many teachers favor the use of portfolios as an assessment tool. Portfolios do a great job of showing students, parents and teachers the progress a student has made over time. However, questions about reliability currently hinder the use of portfolios for assessments where decisions about promotion, retention or graduation are involved (Koretz, Klein, McCaffrey & Stecher, 1993).

The best evidence about a student's learning is collected and analyzed data from both sources: classroom assessments and district level accountability assessments. Once alignment has been established, assessment for accountability will only verify what students and teachers already know from classroom assessment. Table 10 shows how classroom assessment fits in traditional and standards-based systems.

**Table 10**

<b>PRACTICE</b>	<b>FROM</b>	<b>TO</b>
<b>WHAT ROLE DOES CLASSROOM ASSESSMENT PLAY IN THE ASSESSMENT PLAN?</b>	Classroom assessments are selected by each teacher and the scoring systems are unique to each teacher.	Classroom assessments are aligned with district accountability tests and content standards.
<b>IMPACT</b>	<p>The teacher selects any form of assessment.</p> <p>The teachers use their own systems of ranking students.</p> <p>Decisions about promotion, retention, ranking, grades and graduation are assigned to the teacher(s).</p>	<p>A variety of classroom assessments are used, all of which align with accountability tests and standards.</p> <p>Common rating systems are used for evaluation.</p> <p>Decisions about promotion, retention, ranking, grades and graduation follow common formats.</p>

## Programs

Programs are the easiest organizational element to impact but represent the level that has the least impact on student achievement. Programs are those things that can be purchased and implemented as a unit. Curriculum materials and instructional strategies generally fall into this element of the organization. Traditionally, curriculum adoption followed a multi-year cycle. Each content area waited its turn to go through the adoption process. Money was allocated for the purchase of new texts and materials according to that schedule. In between adoptions, teachers would supplement the curriculum with their own materials. Teachers developed favorite units which might or might not fit with the district scope and sequence. In the standards-based system, the development of curriculum and instruction occurs after consensus has been reached on content standards and after the format of accountability assessments has been determined. The purpose of curriculum and instruction is to provide the kinds of experiences that result in learning for each student such that all students meet or exceed performance standards (see Table 11).

**Table 11**

<b>PROGRAM</b>	<b>FROM</b>	<b>TO</b>
<b>HOW IS THE CURRICULUM SELECTED?</b>	Cyclical textbook adoptions become the de facto curriculum.	Content standards and benchmarks determine what is taught.
<b>IMPACT</b>	<p>One textbook series is used throughout the system.</p> <p>One textbook is used in a class.</p> <p>Since textbooks are written for California, New York, Texas and Florida, schools must choose the curriculum framework of those states.</p>	<p>No textbook, one or several texts may be used across the levels. Supplementary materials may be required for some students.</p> <p>Many different resources may be used in a class to meet the specific needs of individual students.</p> <p>Districts have more control over what students learn.</p>

## Procedures

Data management within the assessment system is part of the "procedures" element of the organization. The assessment system is directed by all of the previously described elements. Standards-based systems that use student data in the instructional decision-making process require data management systems that provide timely and accurate data about each student. Data management systems unable to provide data in "real time"--or to handle efficiently student transfers, new students, and students who have left the system--may need updating (see Table 12).

**Table 12**

<b>PROCEDURE</b>	<b>FROM</b>	<b>TO</b>
<b>HOW ARE STUDENT ACHIEVEMENT DATA SCORED?</b>	It is not necessary to have current district data available to teachers because the data do not impact instruction. Teachers keep the data they value to themselves.	All academic data from classroom and district level accountability assessments are important in making instructional decisions for students.
<b>IMPACT</b>	District level data storage can be on paper. No need for retrieval.  No need for teachers to enter or access achievement data.	School and district-wide electronic networks are essential to store and retrieve achievement data.  System must be teacher friendly, easy to enter and access data.

**When to test:**

When to allow students to attempt or "challenge" a test is another "procedure." Standardized norm-referenced assessments require students to be tested in either the fall or the spring "norming window." These tests are designed on the premise of the normal curve where students' scores reflect a range of readiness. School improvement is gauged by increases in the mean or median percentile rank scores. In other words, if the mean student score was higher than the previous year, the school or district is believed to have done a good job.

Assessments that measure students against standards do not require all testing to be done on the same day. Students may formally challenge assessments "on demand" when the student and the teacher believe the student is ready to be successful rather than having all students test at the same time and on the same day. We know that all students are not ready on the same day or during the same week. Students will do much better if they can challenge the test/assessment when they are ready rather than when "we" are ready to test them. "On demand" assessment systems give the "when to test" decision to the teacher and the student (see Table 13).

**Table 13**

<b>PROCEDURE</b>	<b>FROM</b>	<b>TO</b>
<b>WHEN TO TEST?</b>	Accountability tests are given during the norming window in the fall or spring. All students take the test at the same time.	Standards-based tests are offered "on demand" when the teacher and the student believe the student will be successful.
<b>IMPACT</b>	All students test at the same time, ready or not.  Since scope and sequence of the norm-referenced test usually does not match the curriculum, students often test on material they have not studied.	Students test when they are ready.  Since assessments, curriculum and instruction are aligned with content standards, students test on what they have been taught.

**When to score and report:**

Standardized norm-referenced tests are machine scored in the months following test administration. Scoring by the testing services takes approximately four weeks or more. It is not unusual for data to be available to teachers when the students are ready to leave school for the summer or when teachers return in the fall. Local scoring options have reduced the delay in returning data to teachers, but the volumes of paper generated are cumbersome and difficult for teachers to manage.

Standards-based systems that allow on-demand assessment require "real time" scoring systems. Scoring systems must be devised that provide immediate and continuous scoring and delivery of data to students, parents and teachers. Achievement data have the most instructional impact when they are available immediately.

These systems also allow students another chance to meet or exceed performance standards. If students are offered more than one chance, issues around alternative formats and testing intervals need to be resolved (see Table 14).

**Table 14**

<b>PROCEDURE</b>	<b>FROM</b>	<b>TO</b>
<b>WHEN TO SCORE AND REPORT?</b>	Scoring may occur anytime after test administration. Reports are developed the next fall.	Scoring and reporting need to be continuously available to provide students, teachers and parents with current and accurate achievement data.

<b>IMPACT</b>	<p>Scoring is done in bulk and is sent away for scoring by a service company.</p> <p>Reports may be given to the parents and teachers at the end of the school year or the following year.</p>	<p>Scoring systems must be developed locally to provide quick and accurate service and must automatically download into individual student data bases.</p> <p>Reports are returned to the students, parents and teachers within days of the challenge--in "real time."</p>
---------------	--	--

### Reporting formats:

Teachers, students and parents want reporting formats that are easy to understand. They do not want reports they have to read and analyze. Standardized norm-referenced tests have met this need by use of one chart formats comparing the student's achievement to the national sample of other students on each of the basic skill areas. Similarly, new assessments need to provide parents with crisp and clear messages about what their children know, what they can do, and what they need to learn. Standards-based systems using combinations of text and graphic formats are easier for students, parents and teachers to understand (see Table 15).

**Table 15**

<b>PROCEDURE</b>	<b>FROM</b>	<b>TO</b>
<b>HOW ARE REPORTS FORMATTED?</b>	A single chart that compares the student's score with national sample on basic skills.	An individual report for each content standard that compares the student's accomplishment against a performance standard.
<b>IMPACT</b>	<p>One chart covers all basic skill areas.</p> <p>Student scores are compared to a norming sample.</p>	<p>One report for each content standard.</p> <p>Student accomplishments are compared to a performance standard.</p>

---

## Conclusion

The classroom, school, district and community are all interacting organizational systems. Within each system, organizational elements direct how the organization functions. Purpose is the core element. It is the most difficult to change and is protected by the other elements-principles, policies, processes, practices, programs, and procedures. However, changing the more central elements has the most dramatic and sustained impact on student achievement. More distant elements-processes, practices, programs and procedures-are much easier to change but have the least sustained impact on the system.

Standards present educators with an opportunity to make changes in the core elements of the educational system. However, initiatives that attempt to change the purpose, principles and policies of the organization will be purged unless the individuals in the system can be taken through a change process. Changing the culture of the school or district requires a commitment of time and energy to the change process. There is no magic bullet. There is no package that can be purchased that has sustained systemic impact. Teachers, administrators, parents, and other members of the school community have to value public education enough to be willing to do the hard work in the process of change. Only then will sustained high achievement for all students be attained.

## References

Baker, E. L., & Linn, R. L. (1993/94, Winter). Towards an understanding of performance standards. *The CRESST Line*, 1-2.

Burger, S. E., & Burger, D. L. (1994). Determine the validity of performance-based assessment. *Educational Measurement: Issues and Practices*, 13(1), 9-15.

Challenging technical news for performance assessment. (1993, Spring). *The CRESST Line*, 6-7.

Cordell, F. D., & Waters, J. T. (1993). *Improving student performance: New strategies for implementing higher standards*. Greeley, CO: The Center for Peak Performing Schools.

Covey, S. R. (1991). *Principle centered leadership*. New York: Simon & Schuster.

Guskey, T. R. (1994). *High stakes performance assessment*. Thousand Oaks, CA: Corwin.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Hymes, D. L., Chafin, A. E., & Gonder, P. (1991). *The challenging face of testing and assessment, problems and solutions*. Arlington, VA: American Association of School Administrators.

Kendall, J. S., & Marzano, R. J. (1995). *The systematic identification and articulation of*

*content standards and benchmarks: Update*. Aurora, CO: Mid-continent Regional Educational Laboratory.

Koretz, D., Klein, S., McCaffrey, D., & Stecher, B. (1993). *Interim report: The reliability of the Vermont portfolio scores in the 1992 school year* (CSE Tech. Rep. 370). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Merhens, W. A., & Popham, W. J. (1992). How to evaluate the legal defensibility of high stakes tests. *Applied Measurement in Education*, 5(3), 265-283.

Phillips, S. E. (1993). Legal issues in performance assessment. *West's Educational Law Quarterly*, 2(2), 329-358.

September 24, 1997