Urban School District Accountability Systems

December 2003

Table of Contents

List of Tables and Figures

Executive Summary

While much is known about state accountability systems, relatively little is known about district accountability systems. This report, which is based on a survey of district accountability systems, is designed to narrow this knowledge gap. The Center for Reform of School Systems (CRSS) prepared the report as a subcontractor to the Education Commission of the States (ECS), under a grant from the U.S. Department of Education.

The primary objective of this report is to examine to what extent large urban districts have accountability systems in place, to put forth a set of core principles for comprehensive district accountability systems and to interpret the survey results in terms of the principles to create a clearer picture of how accountability systems are currently being implemented. Finally, it is hoped this report will be an effective advocacy document for establishing and improving district accountability systems.

The report is based on a telephone survey of the nation's largest 120 districts. Acceptable responses were obtained from 99 districts (an 83% response rate). Of these, 81 districts administer districtwide tests beyond what the state requires – although one-third of the 81 have their own rating or ranking systems. And while 60 districts reward, sanction or intervene in schools based on school performance, most of them (36) base their actions on results from state accountability ratings, not district ratings.

Based on four broad principles of district accountability and five indicators of accountability, the 99 districts were classified into four categories – threshold, emerging, complementary and state. The "threshold" districts, which have the most comprehensive accountability systems, are Atlanta, Boston, Cincinnati, Clark County (Nevada), Dallas, Houston, Minneapolis, Newark and San Francisco.

This report provides essential information about the accountability systems in these nine cities, with three – Boston, Cincinnati and Houston – the focus of short essays.

Measurement is the first step that leads to control and eventually to improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it.

– H. James Harrington

Introduction

What is Accountability?

Accountability is holding people responsible for meeting standards. Accountability is at the core of standards-based school reform. Without accountability, standards are not really standards, but rather just goals.

Research has shown that an aligned system of content standards, performance standards, assessments and consequences is a powerful force for improving student achievement.[1] Texas, North Carolina, Massachusetts, Maryland and California are among the states that report increasing student test scores on state assessments. Many educators, researchers and policymakers attribute these improved test scores, at least partially, to well-understood accountability systems that include real consequences.

Several urban districts also have put accountability systems into place. In the opinion of district leaders, these accountability systems have contributed significantly to improved student achievement. The Houston Independent School District, for example, established a district accountability system in 1994. Since then, Houston pass rates on the Texas Assessment of Academic Skills (TAAS) have outpaced improvements in other large districts in Texas, and the achievement gap between Houston and the rest of the state has significantly narrowed. In another example, Cincinnati Public Schools, which implemented a district accountability system in 1998, is the only urban district in Ohio that has been taken off the state's academic watch list.

Without question, Houston and Cincinnati have done more than establish district accountability systems. Both districts have implemented comprehensive reform agendas to improve student achievement. District accountability systems, like state accountability systems, must be part of a larger whole.

Purpose of this Report

While much is known about state accountability systems, relatively little is known about district accountability systems. This report, which is based on a survey of 120 districts, is designed to address this knowledge gap.

These districts are important because they disproportionately serve the nation's poorest children. The 120 largest districts – out of over 15,000 districts nationally – enroll about 23% of the nation's public school students and about 40% of the nation's children of color. Over half of the

children in these districts are eligible for free or reduced lunch. [2] Erasing the achievement gap in large urban districts is a national priority.

The Center for Reform of School Systems (CRSS) prepared this report as a subcontractor to the Education Commission of the States (ECS), which has been funded by the U.S. Department of Education.

The primary objective of this study is to discover to what extent large urban districts have accountability systems in place. A second objective is to develop a set of core principles for comprehensive district accountability systems. A third objective is to interpret the survey results in terms of the principles, to show how accountability systems are currently being implemented. Finally, it is hoped this report will be an effective advocacy document for the establishment and improvement of district accountability systems.

The report comprises five parts. They are:

1. Defining district accountability systems to show how they are a necessary part of standards-based reform and linking standards-based reform with comprehensive theories of action for change

2. Proposing four principles to assist policymakers in evaluating existing district accountability systems and developing new ones, which establish the framework for the analysis of existing district accountability systems

3. Reporting the results of a telephone survey of the 120 largest districts in the nation that aided the development of a classification system and placement of the 99 responding districts into one of four categories

4. Providing a more in-depth description and analysis of three districts that have implemented district accountability systems as an essential part of systemwide reform

5. Presenting closing observations for policymakers who wish to develop or improve district accountability systems.

Audience

This report was written primarily for district policymakers: school boards, superintendents and school administrators in district cabinet-level positions. Civic and business leaders also will find it useful by helping to clarify questions they should be asking of district leaders, and by providing a starting point for conversations about district performance. It should be noted that because accountability systems reveal low performance as well as high performance, there are [3]those who will fight against establishing such systems. District leaders need civic support to fully implement accountability. Just as in other areas of district reform, civic capacity for change is important. Researchers also will find this report of interest.

What is a District Accountability System?

The telephone survey revealed that accountability means different things to different people. Many times, survey respondents asked what was meant by accountability. Respondents frequently described accountability systems that ultimately did not hold anyone accountable.

Accountability systems identify important performance indicators, measure performance using these indicators, collect and distribute performance data, and apply predetermined consequences (positive or negative) to those responsible for achieving predefined outcomes. Simply put, accountability is holding people responsible for results. District accountability systems hold schools and school people accountable for student achievement and for other indicators of school performance.

The two key questions for establishing whether or not a district has an accountability system are: (1) Does the district rate or rank schools based on student achievement (and perhaps other factors)? (2) Are there consequences? Rating or ranking schools is a method by which districts can communicate a great deal of information about a school with only one word or one number. Consequences generally fall into three areas: rewards, sanctions and/or interventions.

Inasmuch as many states rate schools and follow through with rewards, sanctions or interventions, schools in these states are being held accountable. But it is the state holding schools accountable, not the district.

District accountability systems go beyond what the state requires. A district may use the state assessment but set higher standards for rating schools; use additional assessments to measure student achievement; or use additional indicators about students, school climate or parent satisfaction to determine whether or not schools are performing satisfactorily. District accountability systems set district standards for rating schools and respond with district consequences.

Why Have District Accountability Systems?

Some argue that a district system is redundant when a state system is in place and that dual systems might confuse parents and the public.

District accountability systems have at least three major advantages over state systems. Districts can:

- Focus on local priorities
- Refine their accountability systems to measure performance in multiple areas in multiple ways
- Adjust local accountability with comparative ease, making continuous improvement possible.

State accountability systems apply equally to all districts in a state, which can be a diverse group even in a small state. It is impossible for state accountability systems to reflect the priorities of each district. State systems paint with a broad brush. They must consider student achievement in core subjects only.

District accountability systems can consider student achievement in additional areas such as science, history, a second language or the arts. They can develop end-of-course examinations to assess high school performance. They can measure in their own ways student attendance, graduation rates, school safety, parent and community satisfaction or other indicators of school performance.

School districts exist, among other reasons, to provide citizens with local control of their schools to meet the needs of their communities. Large urban districts have unique priorities and needs. They need accountability systems tailored to meet these priorities and needs.

District accountability systems provide districts with a second advantage: districts can fine-tune measurements. For example, a district may wish to measure performance by assessing student achievement in multiple areas at every grade level every year, even if its state does not require it to do so.

State accountability systems typically rely on one assessment. A district might want to use two: perhaps a criterion-referenced test (CRT) and a norm-referenced test (NRT). A district may wish to set a higher standard than the state's for acceptable performance or to demand greater improvements from year to year. It may wish to establish multiple cut-points in assessments to measure performance at multiple levels, for example, basic, proficient and advanced. States generally assess student achievement with one blunt instrument. Districts have the opportunity to measure achievement with multiple sharp instruments.

Flexibility is another advantage of a district accountability system. It is difficult for a state to develop and implement a state accountability system. Inevitably, what emerges is a compromise that meets the political needs of the governor and legislature and accommodates as many special interest groups as possible. Once in place, a state accountability system is difficult to change. Politics get in the way and, in any case, stability is important for identifying trends and measuring progress.

Like the states, school districts operate in a political environment and value stability. But districts can more easily fine-tune or expand their systems to close loopholes, raise standards, measure achievement with additional assessments, change weights for accountability measures or in other ways continuously improve their systems.

The No Child Left Behind Act

Some suggest that the No Child Left Behind Act (NCLB), the reauthorization of the Elementary and Secondary Education Act, diminishes the value of a district accountability system. NCLB significantly changes the context in which states and school districts work. After all, one of the four key principles of NCLB, signed into law by President Bush in January 2002, is accountability for results.

NCLB mandates that "states must develop and implement a single, statewide accountability system that will be effective in ensuring all schools make adequate yearly progress, and hold

accountable those that do not."[4] By 2005-06, states must begin administering annual statewide assessments in reading and mathematics for grades 3-8. And in 2007-08, states must implement science assessments. Will the comprehensive state accountability systems required by NCLB make district accountability systems redundant?

Many districts take this position. Almost every district surveyed for this report expressed concern about NCLB's impact, and many said they were taking a "wait and see" approach based on how their state defined "adequate yearly progress." Many district leaders cited NCLB as a reason not to establish a district accountability system.

The authors of this report think this is a mistake. A school district can move quickly to develop and implement a district accountability system that is consistent with NCLB's requirements. Even if the U.S. Department of Education and all the states follow through with comprehensive implementation of NCLB, these state accountability systems will not come close to providing the local focus, sophisticated system of assessments and flexibility available through district accountability systems.

Districts Should Measure What They Value

What gets measured gets done, sometimes at the expense of what is not measured. Because of this, districts should measure what they value. Districts should be interested in student achievement in reading, mathematics, writing and much more. To get this achievement, districts should have accountability systems that measure it.

Most do not. The fact that relatively few districts hold themselves accountable for student achievement beyond what their states require is surprising. Clearly, district policymakers value much more than reading, mathematics and writing at basic levels of performance. Yet if district policymakers value more than the basics, why do they settle for measuring only the basics? Perhaps most district policymakers do not understand the power that district accountability systems have to drive improvements in student achievement.

During survey phone calls, many district officials explained emphatically that they held schools accountable by intervening in schools that their state rated as low performing.

When respondents were asked if the district put additional demands on schools to reach an even higher standard or to perform well on other measures of achievement, many responded, some almost indignantly, that it was not their role to put pressure on schools to perform but rather to provide assistance so schools *could* perform. One district leader explained: "It's a high-stakes environment. We get enough kicks from the state. We need to support [schools]."

Perhaps it is not surprising that some district policymakers prefer to let the state be the "bad guy with the stick" and to position the district as the "good guy" who helps schools learn how to avoid being hit. However well-intentioned, this attitude deprives districts of a powerful tool to improve student achievement.

6

School systems and schools exist to educate students. Their core activity is teaching and learning. How can a school system or a school continuously improve if it does not measure growth in student achievement? As quality management teaches: what you value you measure; what you measure you get. High-performing organizations measure almost everything. It is almost inconceivable that a school system would not want to know the answer to the most fundamental of all questions: are the children learning?

Principles of an Effective District Accountability System

This study was begun with a "chicken-or-egg" problem. Should school districts be surveyed first and then, from information gathered, develop district accountability principles, or should the principles be developed first in order to design appropriate survey questions? It was determined that the principles must come first, while acknowledging that information gathered in the survey might further refine the principles.

Advisory Panel

To assist in developing the principles, a panel of experts was convened. As demonstrated in Appendix A, some of the most knowledgeable researchers, policymakers and practitioners in the nation were chosen. The panel members included:

- Dale Ballou, associate professor of public policy and education, Vanderbilt University
- Arthur Griffin Jr., chairman of the board of education, Charlotte-Mecklenburg Schools
- Brian Jacob, assistant professor of public policy, Kennedy School of Government, Harvard University
- Colin Martin, director of research and accountability, Gwinnett County (Georgia) Public Schools
- Abelardo Saavedra, executive deputy superintendent for school support, Houston Independent School District
- Susan Sclafani, counselor to the U.S. Secretary of Education.

The panel convened at the CRSS offices in Houston, Texas, on August 19-20, 2002. On the first day, panel members discussed and critiqued the project's goals and approach. On the second day, they worked to outline the characteristics of a district accountability system and to develop a method for classifying district accountability systems.

Based on previous knowledge and the discussions with panel members, a set of principles and survey questions were developed. Following the survey, and with the benefit of the information gathered in the survey, four principles for district accountability systems were defined. Panel members contributed immeasurably to this work and their assistance is acknowledged. At the same time, full responsibility for the principles rests with the authors.

Rather than develop a long list of points to describe the essential characteristics of an ideal district accountability system, four broad areas capture the core ideas. The four principles are in fact a grouping of four short essays. Each one starts with a statement that captures the core

principle and then adds sub-principles and explanations to clarify intent and application. It is hoped policymakers will find these principles useful in developing and improving district accountability systems.

Principle One

District accountability systems must be embedded in a comprehensive theory of action that includes all the elements of standards-based reform, school empowerment, and district and school capacity for high performance. And because accountability systems exist to improve teaching and learning, every part of the system must be designed with this purpose in mind and aligned to make it possible.

Accountability is one link in the chain of standards-based reform. And standards-based reform must be firmly embedded in a coherent theory of action for change that includes empowerment and capacity for high performance.

Standards-based reform is not just tests and consequences. It is a chain with five essential links: content standards, performance standards, aligned curriculum, assessments and consequences. With any of these links missing, the chain fails. These few words capture a huge body of theory and practice with numerous areas for disagreement and probably hundreds of technical difficulties.

Standards-based reform is itself part of a larger whole. It must be embedded in a school district's theory of action for change. It must be aligned with numerous policies, systems and practices that collectively make up the operating system called a school district. These policies, systems and practices can be summarized under two heads – empowerment and capacity for excellence.

Principals and teachers cannot fairly be held accountable for student achievement if they do not have maximum control over their work. It is a management axiom: responsibility and authority go together. Lumped under the word "empowerment" are all the policies, systems and practices that enable principals, teachers and others in the system to control, to as large a degree as possible, the environment in which they do their work.

This opens up the issue of how much decentralization is appropriate – what decisions should be made at what level in the organization. The highest-performing urban districts in the country are still working through this issue, but there is broad agreement that principals, who have the greatest responsibility, must have considerably more authority on budget and personnel than most have currently.

A district's central office must build and control management systems, especially financial, personnel and data systems. These are the systems that hold the district together. A district must also build and control curriculum, professional development and assessment systems. A district cannot manage its core business, teaching and learning, if it does not control these systems. Everything else, unless a strong case can be made to the contrary, should be pushed out to the schools.

Empowerment must be matched with capacity for excellence. To be effective in a standards-based environment, teachers need safe classrooms, adequate resources, ongoing professional development and all the support often subsumed under the word "capacity." Clearly, standards are not a substitute for a well-managed organization. Without a high-quality infrastructure that is dependable and supports principals and teachers, high-stakes assessments are unfair.

This means, specifically, that with the establishment of accountability comes the requirement that the district put a comprehensive instructional management system into place. The four essential elements of a comprehensive instructional management system are curriculum, professional development, formative assessments and a comprehensive student-information management system.

If a district expects teachers to teach to standards, it should provide teachers with a curriculum linked to the standards. The curriculum should be comprehensive, complete, accessible and user friendly. Flexibility should be built into delivering the curriculum, and teachers should have maximum freedom to employ various methodologies and materials, depending on their performance.

District administrators should recognize that in most large urban school districts many of the teachers are not experienced veterans teaching in their field of specialty. More likely, a large number of the teachers fall into three groups: (1) those with only a few years of experience, (2) those teaching out of field or (3) permanent substitutes without a certificate. A comprehensive curriculum provides these teachers with much-needed support.

Professional development linked to the curriculum also is required. School districts must make certain that when teachers are assigned to teach a subject, they know the subject and how to teach the subject to children. One of a central office's most important responsibilities is the creation and management of a professional development system linked to the curriculum. All the principles of adult education, principles known and practiced by the human resources departments of America's best companies, apply here.

The assessments of student achievement that are the linchpin of accountability systems are summative. Economists would call these "lagging indicators" of student achievement. In addition to these, school districts should develop and require schools to use formative assessments. Management also may wish to assess key processes related to instruction. These are leading indicators of student achievement. An effective accountability system measures leading and lagging indicators. The lagging indicators feed into the district accountability system. The leading indicators feed into the instructional management system, part of the district's capacity for excellence.

One of the most important parts of the instructional management system is a robust student-information management system. To manage the instruction of grades, classes, groups of students and individual students, principals and teachers must have disaggregated student achievement data, down to the individual student and test item, available on demand. In order to allocate resources, improve professional development, design intervention strategies and coach

teachers and principals, area and central office administrators must be able to analyze comparative and trend data in numerous ways.

Bringing all this theory and practice together to create a high-performing urban school district is not easy, and indeed it has not yet been done. Alignment is difficult. In theory, standards and accountability, empowerment and capacity must be closely aligned and advance as an integrated system. In the real world of implementation, it is not always possible to maintain this alignment. This does not mean, however, that districts cannot do anything until they can do everything.

Accountability in public school systems is a surrogate for Adam Smith's invisible hand. In efficient markets, what enriches the producer also benefits the consumer. Self-interest is important. No system can depend on altruism, important as it is. Accountability gives the adults in a public school district an incentive to behave in ways that benefit children. A comprehensive accountability system builds on this incentive by making it as easy as possible for adults to change their behavior in ways that benefit children.

The first principle is complex, but the point is simple. A district accountability system is part of a complex whole, and it must be embedded in the whole.

Principle Two

Schools are the primary units of accountability. Student achievement must be the dominant measure of school performance, but other metrics count. Schools must be held accountable for the performance of all students (with minimal exemptions) as well as the performance of selected student groups. Changes in performance over time as well as performance levels should be measured. Also, all other functional units of the district should be held accountable for performance.

Because it takes many teachers to teach a child, because policymakers want teachers to work together and not compete with one another, and because schools really are the units in which children learn, schools are the primary units of accountability. This means that schools receive performance ratings, schools as a whole receive rewards and schools are the objects of interventions.

Student achievement is the dominant measure.

Outcome measures are more important than process measures, and student achievement measured by standardized assessments and graduation rates must be the dominant outcome measure. Other measures, however, are important also. Districts may wish to measure student attendance, school safety, school climate, teacher attendance, parent satisfaction or other measures of school performance. Districts may wish to hold high schools accountable for the percentage of students taking a recommended or advanced curriculum. Measuring the percentage of graduates who receive an advanced diploma is an effective way to improve high school achievement.

Normally, outcome measures are part of the accountability system established by the school board, and process measures are part of management's infrastructure for supporting school improvement.

Because multiple metrics contribute to a school's accountability rating and not all metrics are equal, metrics must be weighted. Considerable thought must be given to the attachment of weights to particular metrics. A three-part rule of thumb for productivity measurement is (1) measure too much and the workforce loses focus, (2) measure too little and the workforce will focus on what is measured at the expense of what is not measured and (3) weight a metric below 5% and it will receive little attention. Given this, a district would be wise to use no more than four or five outcome measures, and the weight given to graduation rates must be high enough so that schools have an incentive to improve the performance of failing children rather than flush them out of the system.

An alternative method of combining multiple metrics is to link accountability to the lowest-performance indicator. The lowest-performance indicator method allows policymakers to use as many performance indicators as needed to measure desired learning outcomes. Focus is maintained on all outcomes by counting failure on one outcome as failure on all. An example is a high school exit standard that requires students to achieve a passing grade in all courses taken and also to pass content examinations in multiple areas. Lowest-performance accountability forces a school to focus equally on all performance indicators. Using too many indicators, however, can make it almost impossible for a school to show improvement and thus may become a demotivator.

A school could be held accountable using both methods. Some metrics could be weighted and other priority metrics could be subjected to the lowest-performance standard.

Measure performance of all students.

One unfortunate but inevitable consequence of accountability is that some adults will attempt to "game" the system. Some critics of accountability have used this as a reason not to have accountability. But what is required is simply a system that is difficult to game.

Clearly, all children count, and schools must be accountable for all children. This requires minimal exemptions from assessments and fail-safe test administration. Establishing a test administration system that makes cheating almost impossible is fairly simple. But it is not fair or even possible to test all children. What a district must do is make certain that schools are not encouraging children they have failed to teach to stay home on test day or exempting special education or English language learners who should be tested even though they are at greater risk of failure.

One way to do this is to hold schools accountable for the achievement of all students enrolled rather than the achievement of all students tested. This encourages the school to test every available child. (Special provisions can be made for children with severe learning disabilities or

children who have been in the country for less than a year and cannot be tested in their native language.) The point is that schools should test every child possible. Every child counts.

Schools must be held accountable for the performance of student groups that have historically been neglected by the school system. There are some Americans who believe that poor and minority children have not been neglected by the public school system. They believe these students perform poorly for reasons outside the control of the public schools. It is true that many poor children come to school less ready to read than most middle-class children and that instructional approaches that work best with middle-class children may not work as well with poor children. It is also true that under most circumstances, it costs more to educate a poor child than a middle-class child.

Notwithstanding these points, the authors believe that public schools have neglected poor children. Less is expected of them. Less is provided to them. Too many children have failed, and all Americans must accept responsibility for their failure.

Public attitudes and some state and school district policies have changed in recent years. The No Child Left Behind Act to some extent may make the point moot. It requires that test data be disaggregated by student groups – African American, Hispanic, Native American and poor – and holds schools accountable for the performance of these groups. Also, as in the Texas Accountability System, it uses the lowest-performance accountability standard, so if one group fails to meet the standard, the entire school fails to meet the standard.

District policymakers need not be constrained to disaggregate the data only for groups required by NCLB. For example, a district might identify a neglected student group to add to those whose performance is mandated to be measured by NCLB.

In schools with a low enrollment of minority or poor children, this principle is especially important, for in these schools overall school performance can be excellent even while an entire group of students is being underserved. In schools that are predominately minority and poor – a large percentage of urban schools – this principle is not as important, for the school cannot perform well unless minority and poor children perform well.

If a subpopulation of students is measured as a distinct group, however, the subpopulation should meet a reasonable size threshold. A threshold of five, for example is too small. A school with only five poor children out of 600 could be labeled low performing because two children failed. Indeed, every failure is one failure too many, but as a practical matter, it would not be fair to the school to be labeled low performing if more than 90% of the children performed at a high level and two children failed. Such a standard could become a demotivator and undermine the purpose of the accountability system.

Performance trend and level should be measured.

Districts should measure school performance by trend and level. This is a fundamental principle for all productivity or quality measurement. Improvement trend measures provide the best

motivation for low performers. Benchmarking against the best provides the best motivation for high performers. A good accountability system should measure both.

It is important to emphasize improvement trend or performance level according to the changes that are desired. For example, an emphasis on improvement trend at the expense of performance level is likely to upset high-performing schools. An emphasis on performance level at the expense of improvement trend is likely to demoralize low-performing schools. These are not trivial points. Selecting the wrong mix of measures or setting standards for performance trend or performance level that are too high or too low can lead to significant turmoil.

All functional units should be held accountable.

Though schools are the primary unit of accountability, accountability should reach into every corner of the district. Principals, who bear the brunt of school accountability, will demand accountability from teachers and more control over their schools. They will demand that functional departments be held accountable for measured indicators of performance. Textbooks must be available when school starts; transportation systems must deliver students to school on time; broken equipment must be repaired quickly; everything must work. If accountability does not spread down into the classroom and up into the central office, schools cannot fairly be held accountable.

Principle Three

All students should be assessed at every grade level every year by standardized tests that are aligned as closely as possible with the curriculum, cover as much of the curriculum as possible, and validly and reliably measure what children know. Also, assessments must be fair and seen to be fair.

Assessments at a few grades are valuable, but assessments at every grade every year are much more valuable. Without assessments at every grade, it is difficult to manage instruction and push accountability into the classroom.

Because the high school curriculum is built on courses rather than grade levels, high school performance should be assessed both by end-of-course examinations and by a high school exit examination. Exit examinations are useful, but it is difficult to link them directly to instruction. Developing end-of-course examinations takes time, so most high school accountability will begin with just an exit examination, but end-of-course examinations should be phased in as quickly as possible.

Assessments must be standardized. Nonstandardized assessments, especially in a high-stakes environment, are not reliable. Too many failing urban schools and rampant grade inflation are proof enough of that.

Though test makers sometimes make mistakes, and continuous improvement of test construction is needed, the science of multiple-choice test construction is well developed. A well-constructed,

standardized multiple-choice test is valid, reliable and fair. Subjective evaluation is extremely important but serves another purpose.

Assessments should, of course, be aligned as closely as possible with the curriculum. It makes no sense to teach one thing and test another, and assessments should cover as much of the curriculum as possible. Reading, writing and mathematics in elementary and middle school and exit examinations in high school are a good place to start. But science and history in elementary school and end-of-course examinations for all core high school courses should follow as quickly as possible. Some districts may also wish to assess foreign languages, the arts or other subjects that are community priorities and have been placed in the curriculum.

Because assessments must validly and reliably measure what children know, and because they must be fair and be seen as fair, multiple measures of the same curriculum are preferred. One measure cannot describe a complex reality. One reason sports are liked is because games simplify life. The team with the most points wins. The health or the performance of a company, however, cannot be measured with one number. Likewise, what a child knows or how well a school is performing cannot be accurately measured with only one form of assessment.

Multiple assessments are not preferred to compensate for poorly designed tests. Poorly designed tests should be redesigned. Multiple assessments are preferred because they measure performance in different ways. Two or more assessments not only more accurately measure what a child knows, they provide backup in case a child has a bad day, and they reassure parents and the public that students have more than one chance to demonstrate what they know.

Multiple measures are not practical for high school end-of-course examinations but are useful for assessing elementary school grade-level performance and middle school basic-skills performance. A prime example of multiple measures is the use of both a state-mandated criterion-referenced test (CRT) linked to the state standards and a norm-referenced test (NRT), such as the Stanford 9, to measure grade level performance in reading or mathematics in elementary school.

Multiple cut-points on a test also are preferred. Multiple cut-points make possible multiple measures of performance on the same test, for example, basic, proficient and advanced. A high-stakes test that measures only basic achievement gives teachers an incentive to focus on children near the bubble and provides little incentive to improve the performance of children already performing in the first or second quartile.

Cut-points for various performance levels on tests must be reasonable. Cut-points that show 70 to 80% of students failing to meet basic standards demoralize students and teachers. Cut-points can be raised as student achievement improves. This means that tests must be constructed so that, by raising cut-points, they can be made more difficult over time. The alternative is to frequently change test series. Eventually, all test series have to be replaced, but the longer a test series can be used, the better. Without a stable test series, it is difficult to measure improvement over time.

A common sense rule prevails when determining when, over what and how children will be tested; and in developing assessments and setting performance standards, the assessment system must be fair, not only to the children but also to educators, and it must be seen to be fair. Implementing an accountability system under the best of circumstances is difficult. Policymakers must bend over backwards to demonstrate fairness, even if this means phasing in the system more slowly than desired. An assessment system that is perceived to be unfair will generate additional resistance and perhaps sufficient public backlash to derail its implementation.

Principle Four

Accountability means consequences, both positive and negative, for schools and for everyone in the system. Everything about the accountability system – structure, process, information about assessments, assessment results, accountability ratings, consequences and more – should be broadly communicated in easily understood language to all district employees, parents and the public.

Without consequences, there is no accountability. One does not measure just to obtain information. One measures to change behavior. Low-stakes tests have some value, but high-stakes tests, even though they may make some students nervous and some educators uncomfortable, are required. The prevailing culture of most urban districts is based on compliance. Consequences for performance combined with empowerment and capacity for excellence are the foundation for a culture based on performance.

What are the stakes? The first consequence for high or low student achievement and other measures of school performance is a school grade. Whatever the criteria and whatever the performance indicator – a letter grade, a numerical grade, a label or something else – schools must be given a rating and ranked. Real accountability requires that school performance be made public and that there be positive and negative consequences for this performance.

The public spotlight is the first and most effective consequence. Educators are among America's most altruistic professionals. Most take great pride in their work. Bold newspaper headlines, Web postings, newsletters to parents and a sign in front of a school proclaiming the school exemplary or recognized motivates teachers, principals and other district employees. Also, the spotlight likely calls forth praise or demands for improvement from parents, attracts the attention of real estate brokers and even affects property values. The spotlight puts even more pressure on school people to strive for excellence.

In addition to the spotlight, districts have at their disposal a wide range of responses to high or low performance. High performance can be rewarded with additional flexibility in school management, additional resources, or group awards or gifts to school employees. Negative responses can range from mandated school improvement plans to audits, district assistance, intervention teams, employee transfers, and partial or full reconstitution.

What applies to schools also applies to other functional units of the district. Transportation, security, facilities management, food service, and central office functions can be rewarded or

disciplined for their success or failure to meet predetermined performance goals designed to enhance their service to schools.

In addition to group accountability, district administrators, principals, teachers, students and, to the extent possible, all others involved in the enterprise must be accountable. Administrators and teachers will respond positively to awards, certificates, events, and other forms of recognition, but cash bonuses should be considered. To have an effect, the amount should be significant.

If possible, student performance at the school and classroom levels should be reflected in principal and teacher evaluations, and have some impact on compensation. Given the single-salary schedule in state education codes and union contracts, this is difficult to achieve. But districts committed to accountability should push for change as the political environment permits. In the long run, a performance culture is incompatible with a salary schedule that rewards teachers and administrators just for time behind the desk or in front of the class. Tenure for principals and teachers also is incompatible with a performance culture. To the extent possible, districts need to be able to assign principals to the schools that need them most or move them to nonleadership positions or out of the system if they cannot perform. The same is true of teachers. Principals must be able to build a team that best meets the needs of the school. There is no place in an accountability culture for teacher job placement by seniority. An effective district accountability system inevitably leads to major changes in human resource management.

The comments above about principals and teachers apply also to area and central office administrators. In fact, every job in the system can be tied to performance indicators, and every school district employee should be subject to similar positive or negative consequences for performance.

Students also must be accountable. This means promotion and graduation standards. This is really nothing new. There always have been promotion and graduation standards: teachers determine whether or not a child is performing on grade level or has passed or not passed a course. The difference is that in a district accountability system, the district includes standardized assessments in the standards. The standardized assessments can be part of a weighted or lowest-performance indicator promotion or graduation standard.

Everything about the accountability system – structure, process, information about assessments, assessment results, accountability ratings, consequences and more – should be broadly communicated in easily understood language to all district employees, parents and the public. This does not mean that everything about the accountability system must be simple; in fact, it is almost impossible to create a simple system. It means the district must develop an effective communication plan to continuously explain, clarify and update the workforce, students, parents and community about the system.

This point brings us full circle in establishing principles. A district develops a district accountability system for no other reason than to improve student achievement. Everything about the system is designed with this purpose in mind. This means everyone involved in improving

student achievement – and that includes the community – needs to understand just what is being done and why.

Methodology and Results

From an earlier draft of these principles of a district accountability system, a set of survey questions was developed for the nation's 120 largest school districts. The purpose of the survey was to discover to what extent districts have accountability systems. Using the principles, the authors also identified accountability indicators. Using the accountability indicators, they classified districts. Nine "Threshold Accountability Districts" were identified as a result of the sorting. Summaries of those systems are included in this report. Also included are three district accountability case studies for Boston, Cincinnati and Houston.

Survey

The survey was designed to collect information about accountability in 120 districts. We also wanted to elicit responses on district climate and attitude towards accountability and the impact of state accountability systems on districts. Generally, the survey includes questions on whether the district rates or ranks schools, what elements are used in rating or ranking schools, whether and how rewards and interventions are used, and whether the district plans any changes to the system. Table 1 shows the linkages between the survey items and the principles. Appendix B contains the survey instrument. Complete responses to the questions are contained in the Accountability Database.

Table 1: Principles of Accountability and Survey Items

Principle of Accountability	Survey Item
1. District accountability systems must be embedded in a comprehensive theory of action that includes all the elements of standards-based reform, school empowerment, and district and school capacity for high performance. And because accountability systems exist to improve teaching and learning, every part of the system must be designed with this purpose in mind and aligned to make it possible.	• Are there specific ways accountability data are used to drive change (improve performance)? • How are performance/progress data collected and tracked?
2. Schools are the primary units of accountability. Student achievement must be the dominant measure of school performance, but other metrics should count. Schools must be held accountable for the performance of all students (with minimal exemptions) as well as the performance of selected student groups. Changes in performance over time as well as performance levels should be measured. Also, all other functional units of the district should be held accountable for performance.	• Does the district rate and/or rank schools based on student performance, attendance, or other criteria? • If a district rates or ranks, what criteria are used to determine a school's rating or ranking (i.e., standardized test scores, growth in standardized test scores, attendance, discipline reports, etc.)? • Are these criteria weighted in determining a school's rating? • Does the district compare a school's current performance data to an absolute performance standard? • Does the district examine a school's overall performance growth over time? • Does the district's accountability system rely on a combination of absolute performance and performance gains?
3. All students should be assessed at every grade level every year by standardized tests that are aligned as closely as possible with the curriculum, cover as much of the curriculum as possible, and validly and reliably measure what children know. Also, assessments must be fair and seen to be fair.	• What tests does the district administer? In what grades? What subjects? How frequent?
4. Accountability means consequences, both positive and negative, for schools and for everyone in the system. Everything about the accountability system – structure, process, information about assessments, assessment results, accountability ratings, consequences and more – should be broadly communicated in easily understood language to all district employees, parents and the public.	• Does the district rate or rank schools based on student performance, attendance, or other criteria? • What are the rating or grade categories that the district assigns to schools? • How are ratings/grades/levels determined? • How are ratings publicized? • Are schools eligible for rewards and/or interventions as part of the accountability system?

Selection of Districts

Once the survey was developed, the pool of districts to be surveyed was identified. The largest 120 districts in the nation, as identified by the National Center for Education Statistics Common Core of Data statistics for 2000-01 were selected. By selecting 120, the authors ensured the large countywide districts in Florida, Maryland, Virginia and Georgia – not all of them predominately urban – and many smaller, yet decidedly urban, districts such as St. Louis, Indianapolis and

Pittsburgh were included. These 120 districts educated 11 million children in 2000-01, or roughly 23% of the nation's students. The 120 districts surveyed are listed along with enrollment information in Appendix C.

Telephone Surveys

The authors of this report determined that telephone surveys would elicit the highest response rate. Phone surveys provide the opportunity to follow up on ambiguous responses, ask for other district contacts, and get a sense of a district's commitment to and enthusiasm for accountability. Telephone calls were the only way to navigate district organizational structures, find the appropriate person to interview and obtain information from more than one person.

Surveyors discovered that information about district accountability systems is difficult to obtain. It took an estimated 850 phone calls to generate 99 completed surveys. For many districts, it was hard – or impossible – to penetrate the bureaucracy from afar. Surveyors encountered some skepticism about the purpose of the survey and its emphasis on accountability. Several districts formally declined to participate.

For a handful of districts (about 10% of the sample), surveyors found themselves in very animated conversations with people who were dedicated to their work. These people were committed to accountability, wanted to learn more about recent accountability activity and wanted to receive the results of this research. Most of these districts are developing student-information management systems that will provide accurate and timely data to principals and teachers about student achievement.

Survey Limitations

Though the advisory panel reviewed the survey, the survey instrument was not tested for validity, nor was it formally piloted. While every effort was made to interview appropriate district personnel (e.g., directors of research, evaluation, assessment or accountability), this was not always possible. For these reasons, surveys of some districts may not be accurate or complete. In addition, because the results are based on the responses of districts that actively chose to participate, it cannot be said that the responses fully represent the nation's 120 largest districts. Eighteen districts did not respond to repeated requests to complete the survey, and three districts declined to participate.

Key Findings

Ninety-nine districts (an 83% response rate) provided acceptable responses. Of these, 81 districts administer districtwide tests beyond what the state requires. See Appendix D for the tests these districts administer. Only 27 districts, however, have their own district rating or ranking system. Sixty districts, according to this survey, reward, sanction or intervene in schools based on school performance. Thirty-eight districts do this based on state accountability ratings, not district ratings. These districts are doing exactly what state policymakers hoped they would do, but they are not going beyond what the state requires. Table 2 provides a summary of information gathered by the survey.

Table 2: Summaries of Survey Responses (N = 99)

Survey Item	Description of Responses	Summary
Does the district administer tests beyond what the state requires?	Details the type of tests the district uses for a variety of purposes and the grades and subjects tested. Often these tests are used for diagnostic purposes and program evaluation, not accountability.	Most districts (81) test beyond what the state requires.
Does the district rate or rank schools?	Determines whether the district rates or ranks schools.	Less than one-third of districts (27) rate or rank schools.
What criteria are used to determine a school's rating or ranking?	Shows whether standardized test scores, growth in standardized test scores, attendance, graduation rates, discipline reports or other factors are used in determining ratings.	Of those that rate, all use student performance measures.
Are these criteria weighted in determining a school's ranking?	Determines whether student performance is the primary measure of performance. If there is a weighting system, a description follows that gives the weight of each indicator.	Of those that rate, student performance is the primary if not sole indicator.
Are ratings based upon performance level or improvement trend or both?	Describes the way districts assess the performance of schools in relation to a set standard and how much the school has improved.	Eight districts base ratings on performance level, 4 on improvement trend, and 15 on both.
What are the rating categories or what is the index?	Lists the various categories in which schools are classified such as exemplary, recognized, acceptable, low performing or the range of an index, such as 1-100.	Majority of districts use categories. Some districts use a numerical index.
How are ratings determined?	Describes in detail the process involved in calculating school performance categories or index scores.	Districts use equations of varying complexity to calculate school performance.
How many years has the district accountability system been in place?	Shows the length of time accountability systems have been in place. New York City has had a ranking system based on reading scores for 32 years – this response was an extreme outlier.	Ranges: 1.5-10 Median: 5 Mean: 6
How are ratings publicized?	Shows how school performance information (as evaluated by the state or the district) is shared with the public.	Most districts use local newspapers and district Web sites to publicize school ratings.
Does the district administer interventions, rewards or sanctions (I/R/S) based on state or district ratings?	Describes the system of school rewards, interventions and sanctions based on district or state ratings.	Most districts (59) provide I/R/S to schools due to state or district ratings.
How are performance data collected and tracked?	Describes district data-collection practices.	Districts describe data management with varying degrees of complexity.
Are there specific ways accountability data are used to drive change?	Describes any additional components of accountability used within the district.	Most often responders conveyed the district's general attitudes toward accountability.
Are there other elements of note, including future changes to the system?	Shows any changes or plans the district has for the future.	This item most often elicited responses relating to No Child Left Behind.

Major findings, based on information in the table, are as follows:

- Most districts (82%) administer districtwide tests beyond what the state requires. Districts use the additional tests primarily for diagnostic purposes. It appears, however, that some districts administer some tests for *no clear reason*. The test results are not collected or tracked in a way the survey respondent either knew about or could easily describe.

- Twenty-seven districts rate or rank schools.

    - Twenty-two of these districts respond to school performance with interventions, rewards or sanctions (I/R/S).
    - Of the five districts that do not provide I/R/S:
        - Gwinnett County (Georgia) uses ratings primarily for internal planning.
        - Tucson and Saint Paul are in the process of adding a system of I/R/S to their accountability systems.
        - Knox County shines the public spotlight on school rankings but uses no additional I/R/S.
        - Conclusive information was not available for Philadelphia.[5]

- Sixty percent of districts provide support for under-performing schools as identified in state or district ratings. This support ranges from reactionary and program-driven interventions to highly targeted support that is strategic and anticipatory. Examples of typical interventions include:
    - Targeted assistance with instruction, facilities or budget management
    - Schoolwide assistance provided by a district team deployed to work with the school staff on instruction, facilities, budget, professional development and data analysis
    - Grant-directed program-specific assistance
    - Data-driven assistance, including training for principals and teachers to use diagnostic data to identify and improve weak spots.

Attitudes Toward Accountability

The survey revealed a mostly negative attitude by districts toward state and federal accountability requirements. A large majority of respondents expressed uncertainty and anxiety about the new testing and accountability environment created by No Child Left Behind. Many districts view their role as supportive of schools. As one district administrator commented, "The district's role is to ask schools what assistance they need, generate reports, and discuss report results with schools and teachers." A third district administrator put it this way, "The state has more or less usurped this process and we are now in a reactive mode." These districts, and many others, seem to view No Child Left Behind as adding a layer of requirements and consequences that will make life just that much more difficult for district administrators and schools.

Most districts in states with well-publicized, relatively long-standing accountability systems considered designing and implementing a separate district accountability system to be redundant. One district administrator commented, "The state does it [accountability]. It's foolishness to have a separate rating system." In districts like this one, accountability departments, where they exist, exist to facilitate the district's compliance with the state's accountability system, collect data and issue reports.

Districts like these frequently supplement the state's response to performance with additional rewards for high-performing schools and additional support for low-performing schools. Some of these districts set district goals or require schools to set performance goals but do not respond with consequences if the goals are not met. The state goals trump the district goals. One California district administrator commented, "We don't levy sanctions if schools don't meet the [district] goals in their academic improvement plans. The state goals are the ones that count. The real target is the state API."[6]

Another issue revealed through the interviews is that accountability means many different things to district administrators. Many administrators spoke at length about accountability in their districts. It became apparent, however, that district administrators use "accountability" to describe many policies and practices. For instance, when surveyors initially asked about accountability, many administrators responded with information on how they supported schools the state system rated as low performing. Another common initial response to questions about accountability was an extended description of district testing practices. For many districts, accountability means testing, even if testing is only for diagnostic purposes.

Accountability Indicators

Based on the accountability principles explained earlier, five accountability indicators were identified to determine whether a district was applying a principle. These indicators include:

1. *Rating or ranking of schools*
   Rating or ranking of schools fits with Principles Two and Four. School ratings convey a tremendous amount of information in one or two words. A rating of "exemplary," for example, conveys a school's status in a concise, clear way that a school report card cannot. School report cards typically provide detailed information in a three-to-five page report that takes time to understand and is not easily summarized. When only report cards are used, it is difficult to compare one school with another and impossible to place schools in rank order.

   Report cards are informative, useful tools for parents and communities, but they are no substitute for a rating and ranking of schools. A rating and ranking system enables a district and the public it serves to see in one number, letter or word, the sum of all the performance information about a school, measured according to a metric that reflects a district's priorities. A rating and ranking system simplifies reality, but without rating and ranking, clear communication and accountability for results is impossible.

2. *Multiple assessments of achievement*
   As stated in Principle Three, multiple measures of student achievement measure school performance more accurately than a single measure. The obvious place to start is with a CRT based on strong, high-quality state or district content and performance standards. If this is not available, an NRT may be a cost-effective interim assessment. The state of California, for

example, used the SAT-9 for accountability purposes until the state's California Standards Tests were developed.

Many districts use an already developed state CRT for their accountability systems. This has many advantages. The state CRT is aligned with the state standards, does not require additional testing of students, and it is free. If a state NRT is available, some districts use it to supplement the state CRT. Other districts purchase an off-the-shelf NRT to supplement the state CRT.

Measures of student achievement include more than CRTs and NRTs. Also important, for example, are end-of-course examinations, graduation rates and enrollment in advanced courses.

3. *Performance level and improvement trend*
   As described in Principle Two, districts should measure performance level and improvement trend. Accountability systems need the buy-in of stakeholders. Both high-performing and low-performing schools need incentives to either maintain or improve performance. For example, if the district's accountability system requires schools to reach a 75% passing standard to receive an acceptable rating, some low-performing schools may be so far below the 75% bar that under even the best conditions they will need three to four years to reach the goal. Schools in this situation have a greater incentive to improve if rewards are given for improvement, even if it is short of acceptable performance. Likewise, it is not reasonable to judge a school that is already high performing by year-to-year improvement. These schools benefit from a system that rewards high performance even if performance drops a few percentage points from one year to the next.

4. *Additional performance indicators*
   As stated in Principle Two, student achievement must be the primary measure of school performance, but additional measures are important. Districts should select performance indicators that reflect district priorities. Examples might include teacher turnover rates, school safety or parent surveys of school climate.

5. *Interventions, rewards or sanctions*
   Consequences drive behavior. The spotlight on school performance alone will drive changes in behavior. But additional consequences are necessary, as described in Principle Four. The consequences are summarized with the three words "interventions, rewards or sanctions." Although each of these three responses is conceived of being implemented differently, their common goal is to change behavior; therefore, they are grouped together.

Table 3 includes all 27 districts that rate or rank schools, and shows the other accountability indicators in place in each district. See Appendix E for a list of all the districts with indicators.

Table 3: District Accountability Indicators

District	Rates or ranks schools	Multiple assessments	Additional indicators	Performance level and improvement trend	Administers I/R/S
Atlanta Public Schools, Atlanta, GA	✓	✓	✓	✓	✓
Boston Public Schools, Boston, MA	✓	✓	✓	✓	✓
Cincinnati Public Schools, Cincinnati, OH	✓	✓	✓	✓	✓
Clark County School District*, Las Vegas, NV	✓	✓	✓	✓	✓
Dallas Independent School District, Dallas, TX	✓	✓	✓	✓	✓
Houston Independent School District, Houston, TX	✓	✓	✓	✓	✓
Minneapolis Public Schools, Minneapolis, MN	✓	✓	✓	✓	✓
Newark Public Schools, Newark, NJ	✓	✓	✓	✓	✓
San Francisco Unified School District, San Francisco, CA	✓	✓	✓	✓	✓
Brevard County School District, Viera, FL	✓	✓	✓		✓
Brownsville Independent School District, Brownsville, TX	✓		✓		✓
Chicago Public Schools, Chicago, IL	✓				✓
Cobb County Public Schools, Marietta, GA	✓		✓		✓
District of Columbia Public Schools, Washington, DC	✓		✓		✓
Fairfax County Public Schools, Fairfax, VA	✓	✓		✓	✓
Garland Independent School District, Garland, TX	✓	✓		✓	✓
Gwinnett County Public Schools, Lawrenceville, GA	✓	✓	✓	✓	
Hamilton County School District, Chattanooga, TN	✓				✓
Indianapolis Public Schools**, Indianapolis, IN	✓		✓	✓	✓
Knox County School District, Knoxville, TN	✓				
New York City Public Schools, New York City, NY	✓				✓
Saint Paul Public Schools*, Saint Paul, MN	✓	✓	✓	✓	
San Antonio Independent School District, San Antonio, TX	✓				✓
School District of Philadelphia**, Philadelphia, PA	✓		✓		n/a
Tucson Unified School District, Tucson, AZ	✓	✓		✓	
Tulsa Public Schools, Tulsa, OK	✓	✓	✓		✓
Wichita Public Schools, Wichita, KS	✓	✓			✓

* Not yet implemented

** No longer in use

Classification of Districts

Using the five major accountability indicators described above, the authors developed a classification system and placed the 99 districts in one of the four groups – Threshold

Accountability Districts, Emerging Accountability Districts, Complementary State Accountability Districts or State Accountability Districts (Appendix F shows full listing). Twenty-seven districts were sorted in either the Threshold or Emerging category (for details concerning the calculations of these districts' ratings, see Appendix G).

Threshold Accountability Districts
The nine Threshold Accountability Districts have the complete package, though sometimes just barely. These districts rate school performance using multiple assessments of student achievement and multiple indicators of school performance. These districts respond to performance with consequences that may include assistance or intervention for low-performing schools and rewards for high-performing schools. Appendix H contains more specific information on these nine districts.

Emerging Accountability Districts
The 18 Emerging Accountability Districts rate or rank schools, but do not respond to ratings with interventions, rewards or sanctions. These districts do not shine the public spotlight on performance or respond with consequences. Instead, ratings are used for one of two reasons: internal planning or to identify low-performing schools. Districts that use school ratings for internal-planning purposes do not make the results public. They use the ratings for performance evaluations, curriculum planning and resource allocation.

Districts in the second group use student achievement and other measures to identify low-performing schools. The rankings are generally not publicized. The rankings are used to determine where financial, curriculum and professional development assistance is most needed. A school's rank does not provide substantial information about the actual performance of the school; it only reveals how well a school is performing relative to other schools in the district. In an effort to compare similar schools to one another, some of these districts adjust performance based on demographic and other characteristics.

Complementary State Accountability Districts
Thirty-eight of the 72 remaining districts perceive their role as responding to the needs of schools the state has identified as low performing and incorporate components of accountability in a number of ways. The authors called these districts are Complementary State Accountability Districts. Some of these districts set annual performance goals for assessments, attendance and course taking without any formalized system of consequences for achieving or not achieving goals.

For example, one district set a goal stating "Each elementary school will reduce the number of 3rd-grade children scoring in the bottom national quarter on the reading composite section of the norm-referenced assessment." Similar goals are set for middle and high school. The district accountability report lists the schools that have or have not achieved the goals and publishes the school reports that detail areas for improvement. Other than the publication of the district accountability report, which shined a spotlight on the schools that do not achieve the district's goals, there are no formalized consequences.

Other districts require schools to produce school improvement plans that set goals for improvement on test scores, attendance or other criteria. District participation in these school improvement plans ranges from virtually no oversight to close tracking of district-defined school-improvement plan indicators. Often, these improvement plans are based on performance on the state assessments. For example, some districts ask or encourage schools to develop school improvement plans that respond to weaknesses identified by the state test. Other districts send district staff to schools to walk them through the process of writing a school improvement plan.

Some districts use test data for program evaluation. "We might identify a need at secondary level for reading comprehension and implement an intervention program; then we will use assessment data and other information to evaluate the program," said one respondent.

Other districts rely on state tests for accountability purposes and develop district tests for diagnostic purposes. Districts like this exist in California, Florida, North Carolina and Texas – all states with strong accountability systems.

State Accountability Districts
Thirty-four districts rely completely on their states for accountability. The authors called these districts State Accountability Districts. Everything – testing, evaluation, reporting and administering consequences – is done by the state or by the district following state mandates. These districts may test beyond what the state requires, but they do not use these tests for accountability purposes. Most of these districts are in states with strong accountability systems. A few are in states with weak accountability systems.

At most, what these districts add to the state requirements are explanations to their communities about the state system. A district administrator in Florida, which has a strong accountability system that grades schools A to F, commented, "The district basically tries to explain the state's grading system. For example, we have a number of locations, but the state's definition of grading applies to only certain schools because the other schools don't have two years of data. Part of the district's role is to explain that information to the community."

Case Studies

The following case studies provide a more in-depth description and analysis of three Threshold Accountability Districts – Boston, Cincinnati and Houston.

Boston Public Schools Accountability System: A Theory of Action, A Work in Progress[7]

Executive Summary
Mayor Tom Menino has put himself on the line for school performance: "I want to be judged as your mayor by what happens in the Boston Public Schools [BPS] … and if I fail … then judge me harshly."[8] Voters seem to feel schools have improved, as Tom Menino was re-elected in 1997 and again in 2001. He has worked to gain more control of Boston's schools.

By all accounts, BPS Superintendent Tom Payzant and his team are leading a deep and broad-based school improvement effort, which is yielding results. Boston student scores on the

Massachusetts Comprehensive Assessment of Skills (MCAS), known as one of the hardest in the nation, improved across all races and grades for the year 2002. The district's efforts have attracted millions of dollars of long-term national funding from the Carnegie, Annenberg and National Science Foundations. While many Massachusetts districts have protested the rigorous state assessment, BPS has taken the stance that such high standards will take time to meet, but can serve as the catalyst for dramatic school improvement over time.

The BPS theory of action for improving student achievement is that high-quality instruction is essential. BPS has built its accountability system to reinforce the theory by measuring instructional practice as a key component of measuring school performance. Instead of leaving it to schools to define high-quality instruction and struggle to find or develop new instructional practice on their own, BPS has defined Six Essentials of Whole School Improvement.

This enables the district to organize around a single vision of what it takes to create excellent instruction and to share best practices across schools. This clear definition and measurement of effective instructional practice, combined with a gradual tightening of the links between the pieces of the management system – including the school improvement planning process, professional development, curriculum support, and rewards and consequences – make Boston's system unique.

Narrative
The narrative below describes how the Boston Public Schools accountability system manifests each of the principles of a district accountability system.

Principle One: *District accountability systems must be embedded in a comprehensive theory of action that includes all the elements of standards-based reform, school empowerment, and district and school capacity for high performance. And because accountability systems exist to improve teaching and learning, every part of the system must be designed with this purpose in mind and aligned to make it possible.*

The BPS Focus on Children strategy states: "Our theory of action maintains that if instruction is at a high level and if the conditions in the schools enable good instruction, students will learn. Therefore, it is important to measure the improvement of instruction and school culture." The superintendent, deputies, chief operating officer and school review teams (known as In-Depth Review Teams) all use the same set of rubrics to evaluate school progress on each of the Six Essentials of Whole School Improvement:

1. Use effective instructional practices and create a collaborative school climate to improve student learning.

2. Examine student work and data to drive instruction and professional development.

3. Invest in professional development to improve instruction.

4. Share leadership to sustain instructional improvement.

5. Focus resources to support instructional improvement.

6. Partner with families and community to support student learning.

In the first five years of BPS' reform effort, 1996-2001, these essentials were known as Six Whole School Change Essentials instead of Essentials of School Improvement. The change in moniker is important, because it summarizes the district understanding of its reform progress and of the district role. In the beginning, the issue was change or turnaround of existing practice. In a turnaround situation, waiting for schools to design their own practices is not an option. Now, after five years, most schools can say they have made serious organizational change, initiated new curriculum and instruction techniques in literacy and math, and must now improve on them.

A recent evaluation by Policy Studies Associates finds that schools that implement the Essentials deeply (known as Effective Practice schools) perform better than all other Boston Public Schools in both the 4th and 8th grade, in both literacy and math with similar student populations. Effective Practice schools had more consistent and stable improvement than all other BPS over the four years studied. Perhaps most impressive, Effective Practice schools raised student performance more than all other BPS schools for both the lowest- and the highest-performing students.[9]

District leaders, along with their partner in reform, the Boston Plan for Excellence, took great care in the design of the Essentials.[10] Beginning with research on how poor schools become effective schools, they sought to define the principles of effective instruction, with enough specificity to guide principals and teachers and to allow measurement. They wanted to be clear about the few critical activities, the essentials, on which people should focus. Ellen Guiney, executive director of the Boston Plan for Excellence, recalls, "There are lots of prescriptions out there that tell teachers and principals exactly what to do every day. Superintendent Tom Payzant never wanted to use this approach; he believed that we had to build a new professional culture of teaching in schools that began with teachers owning their students' performance. This is why we asked schools to pick an instructional area and begin by looking together at student work in this area." The Six Essentials reflect the BPS emphasis on building deep instructional capacity.

The rubrics used to define and measure progress against the Six Essentials give them specificity and clarity that distinguish this list from similar ones that fill education reform literature. This specificity also allows the Six Essentials to become the centerpiece of the accountability and support system.

Principle Two: *Schools are the primary units of accountability. Student achievement must be the dominant measure of school performance, but other metrics should count. Schools must be held accountable for the performance of all students (with minimal exemptions) as well as the performance of selected student groups. Changes in performance over time as well as performance levels should be measured. Also, all other functional units of the district should be held accountable for performance.*

In 1997, Massachusetts set 2003 as the year high school students would need to pass the minimum level of proficiency to graduate. BPS had a long way to go. In 1998, 43% of 10th graders passed the MCAS in language arts and only 25% passed in math. Led by Maryellen

Donahue, BPS director of research and assessment, the district set interim performance targets for each school, by grade and by race. Donahue remembers, "In the beginning, we really couldn't know what kind of improvements would be possible for schools. We just knew that we had to focus on continuous improvement and to celebrate steady growth wherever it happened." Setting interim goals by race was the way BPS focused on closing the achievement gap. Donahue reflects, "Everyone always focuses on the huge gaps. Our approach has been to be very clear about performance levels for each racial group, but to focus on the improvement, or the 'closing of the gap.'"

These MCAS targets are combined with other quantifiable outcomes, including suspensions and student attendance, to create the first dimension of the accountability system. School progress against these goals is reported and measured yearly. These yearly measurements are known as the Yearly Check Point (YCP) review. Progress on the yearly targets is combined with an in-depth review that happens once every four years to create an overall accountability rating. Progress has been gradual but substantial, with 60% of 10th graders passing the language arts MCAS in 2001 and 53% passing math, double the number from four years ago.

The quantifiable measures of performance are then combined with school progress on improving instruction. This is the second dimension of the accountability system. With performance levels so low and the potential gains on the MCAS so unclear, Payzant and his team knew they could not wait for student performance results to tell them whether schools were making the right kinds of changes to improve instruction. Rubrics of the Six Essentials allow schools to conduct self-assessments; In-Depth Review Teams use the rubrics to conduct external assessments as well.

The rubrics used to measure these essentials make expectations much more concrete. Like the accountability system overall, these rubrics have evolved over time as schools have worked to make them more succinct and specific. The rubrics are used for schools to conduct self-assessment, by BPS district leaders to assess school progress, and by the In-Depth Review Teams that conduct thorough reviews of schools every four years for use in the accountability system.

Each Essential has a leading indicator. The leading indicator represents the point at which student performance will improve. That is, if a school can get to a high level of implementation on this indicator, then they are likely to improve student performance. For the first Essential – Use Effective Instructional Practices and Create a Collaborative School Climate – the rubric measures how fully the school has implemented the district's chosen literacy approach across grades and classrooms. The districtwide literacy approach is Readers' Workshop and Writers' Workshop Literacy Across the Curriculum.

In addition to being specific, the rubrics demand high levels of practice and implementation. BPS expects that schools should implement them at a level three or better. As the introduction to the rubric states, "In most cases, this (level) represents a substantial raising of the bar from previous years – an action appropriate six years into the reform."[11] For example, for Essential 3 – Invest in Professional Development to Improve Instruction – the rubric defines level 3 as:

Most teachers observe the practice of the content coach, demonstrate instruction for the coach, receive feedback and participate in debriefing sessions, and their instructional practices reflect the practices learned from the coach. Most administrators also participate.

As the example shows, the rubric takes the general principle of investing in professional development and makes it clear that schools must be using instructional coaches in literacy and math to improve practice. The rubric goes further to say that teachers must actually open their classrooms to coach observation and use the feedback from the coach to make changes in instruction.

Clearly, collecting this kind of information on school progress requires school visits. In fact, it requires that those visiting and rating schools have enough understanding of schools to be able to evaluate educational practices. For example, evaluating how deeply a school is implementing Readers' Workshop requires knowing the program and how to identify its practice in a classroom.

Like other districts collecting data that requires school visits and observations, Boston has struggled with how to do so cost effectively but still credibly and accurately. One approach would be to use teams composed of district experts. The district experts could reliably assess school practice and might also be in a position to support schools in areas of weakness following the assessment. But outsiders to the system can lend the evaluation teams credibility, because if chosen well, they might be viewed as having less of a stake in the outcome. The BPS approach to the In-Depth Review (IDR) seeks a balance of insider experts and credible outside reviewers. BPS cluster school leaders choose the chairman of each IDR; the chairman then carefully selects a balanced team. The deputy superintendents review the team composition. Teams receive training and practice reviews before heading out to the field.

Boston is still working to ensure the training for review teams is adequate and results in consistent, accurate reviews. "The increasing clarity of the Essentials helps a lot, as do the new rubrics," reflects Maryellen Donahue. "The rubrics can really be our way of giving reviewers the advantage of all of our expert learning about how to focus on the few, most important indicators to gauge effective instructional practice." The ultimate vision is that the process of conducting in-depth reviews will build increased districtwide and community understanding of the elements of effective instruction.

While not part of the district accountability system per se, it is individual accountability that spurred these systemwide changes. As noted, the mayor has staked his reputation on school improvement. In 1995, Mayor Menino hired Tom Payzant, a nationally prominent, respected leader as superintendent. In 1996, the issue of whether Boston should move from a school board appointed by the mayor, established in 1991, back to an elected board came up for a public referendum. Although polls showed that the public favored an elected school board, Menino succeeded in convincing the public to approve the appointed board – leaving the reins of control (and accountability) with him and Tom Payzant. Last year, the appointed school board extended Tom Payzant's contract to 2007.

Principle Three: *All students should be assessed at every grade level every year by standardized test that are aligned as closely as possible with the curriculum, cover as much of the curriculum as possible, and validly and reliably measure what children know. Also, assessments must be fair and seen to be fair.*

BPS uses the SAT-9 in English and math in grades 3, 4 and 5 and the Massachusetts Comprehensive Assessment System (MCAS) in English and math in 4, 8 and 10, and history and science in 8 and 10. Not all grades are tested every year, but the tests are administered annually and the MCAS is aligned with the state standards and curriculum.
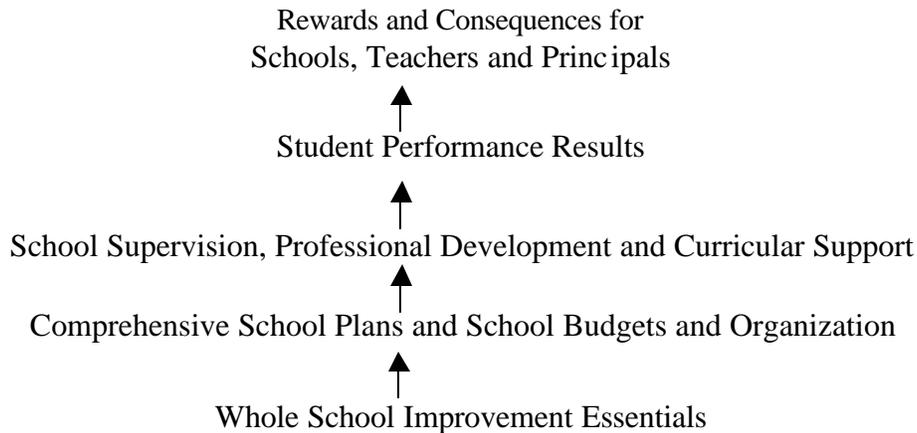
Principle Four: *Accountability means consequences, both positive and negative, for schools and for everyone in the system. Everything about the accountability system – structure, process, information about assessments, assessment results, accountability ratings, consequences and more – should be broadly communicated in easily understood language to all district employees, parents and the public.*

BPS' ratings are clearly communicated to the public through fully annotated reports. The ratings also are available on the district's Web site, and they are published in the newspaper. The following ratings describe whether schools have met performance goals on the Yearly Checkpoint Review: Met or Exceeded yearly progress objectives, Made Substantial Progress toward yearly progress objectives, Made Some Progress toward yearly progress objectives, Maintained performance or Declined. The In-Depth Review process determines whether schools have met the Six Essentials of Learning. The ratings are: Outstanding Accomplishment, Substantial Accomplishment, Some Accomplishment, and Little or No Accomplishment. The combined ratings describe a school's progress toward specific quantitative goals and their progress in implementing the Six Essentials of Learning.

The power of an accountability system lies in whether it motivates teachers and principals to improve practice in ways that improve student performance, and in whether it helps district leaders identify where improvement isn't happening so they can act quickly. BPS hopes that its system motivates improvement because the Essentials give principals and teachers clear direction on the steps to take to improve student performance and provide concrete ways of measuring this. But BPS leaders know the power depends on how clearly principals, teachers and district leaders understand these expectations and on how consistently they are reinforced throughout the system.

Figure 1 below shows the pieces of the system that must be linked. To maximize consistency, the Whole School Essentials is the language and framework for all planning and evaluation throughout the system. BPS has now revised its comprehensive school-planning process to use the same rubrics and format that In-Depth Review Teams use. In this way, schools conduct an annual self-assessment of progress. Schools on the low end of performance receive extra assistance and scrutiny of the self-evaluations in their annual plans. Professional development and curricular support also are increasingly integrated with the Essentials. Each BPS school has an expert literacy and a math coach who work in classrooms with individual teachers and with teams of teachers to implement the district curricular approaches.

Figure 1: BPS System Linkages

Rewards and Consequences for
Schools, Teachers and Principals

↑

Student Performance Results

↑

School Supervision, Professional Development and Curricular Support

↑

Comprehensive School Plans and School Budgets and Organization

↑

Whole School Improvement Essentials

The consequences for success and failure also are becoming clearer. Schools that implement the Essentials most deeply earn a district designation as Effective Practice schools. These schools serve as an enormous resource for the district, sharing their best practices and expertise. They also receive targeted support and participate in their own professional development networks. This year, Effective Practice schools also will receive increased flexibility in their use of professional development resources, receiving a dollar value reflecting the district's combined professional development resources that can be used at their discretion.

Schools receiving a rating on either the yearly checkpoint or the in-depth review of Not Meeting Standards are referred to the superintendent to be supported either through a "targeted change" or an "intervention" process. Targeted change entails a focused, intensive improvement effort supported by the central office's Teaching and Learning Team. Intervention schools are judged to need more far-reaching support. This process begins with an onsite review conducted by an intervention team. The team then makes recommendations to the superintendent that can include steps such as reallocation of current resources, replacement of leadership team, reassignment of staff or addition of specialist resources.

Future Challenges
The district views its accountability system as "a work in progress." As further evaluation and experience pinpoint which of the Essentials have most impact or clarify the practices that matter most, leaders expect to reflect this in the refinement of the accountability rubrics, the evaluation system, and in professional development and curriculum support activities. District leaders also expect to differentiate their support to schools increasingly based on their performance. Current targeted support and intervention efforts will help leaders learn how to more effectively intervene with poorly performing schools. At the same time, the district is working to find more ways of increasing flexibility and opportunity for the highest-performing schools. But the ultimate reward for Effective Practice schools will certainly be the difference it makes for student performance.

Executive Summary

Cincinnati Public Schools (CPS) has devoted great attention to building a meaningful accountability system. The system helps connect the pieces of the district's reform strategy, and it powerfully drives improvements in student achievement by holding both schools and individuals accountable for improvement. To hold schools accountable, the district rates schools based on performance level and improvement trend on the Ohio Proficiency Tests. Schools are subject to rewards and interventions based on their district rating. School's ratings are reflected in principal evaluations, and a new Teacher Evaluation System (TES) links teacher pay more directly to a demonstrated skill in improving student performance. The accountability system was developed as part of an overall reform strategy that includes a sharp focus on improving instructional practice, increasing budget authority at the school level and reconstituting schools that consistently underperform.

Rewards, interventions and instructional improvement are all tied to school achievement. Performance reviews have been redesigned to hold teachers and principals accountable for improving school performance, ensuring that individuals who have the authority and responsibility to deliver high-quality instruction are being held accountable for doing so.

In fall 2002, based on 2001-02 scores, Cincinnati became the first of the six large urban districts in Ohio to move from the state's Academic Emergency list into the Academic Watch category.[13] In 2002, a record number of CPS schools achieved their improvement targets, with 31 out of 72 schools reaching the highest rating category. Of the schools that were designated as the lowest performers four or five years previously, half earned the district's highest rating.[14] Both of the schools that were redesigned in 1998 were rated at the highest-performance level in 2000-01. One of these schools received incentive awards. The accountability system, combined with the support and resources linked to it, is powerfully contributing to gains in student performance.

Narrative

The following narrative describes how the CPS accountability system manifests some parts of each of the principles of a district accountability system.

Principle One: *District accountability systems must be embedded in a comprehensive theory of action that includes all the elements of standards-based reform, school empowerment, and district and school capacity for high performance. And because accountability systems exist to improve teaching and learning, every part of the system must be designed with this purpose in mind and aligned to make it possible.*

The CPS accountability system forms the centerpiece of the district's long-range strategy. CPS has committed significant resources to reward excellence and improve performance. For example, the district implemented a new teacher compensation structure that rewards teacher expertise as measured by student achievement. The district also created team-based schools in which teacher teams have significant control over budget and instructional decisions. Though these two major reform initiatives have gained national attention, it is the coherence of the overall CPS strategy and the commitment of substantial resources that distinguish CPS efforts from the reform initiatives in most other districts.

CPS is now in the sixth year of its long-range strategic plan, Students First. Students First has four critical components:

1. Defining high academic standards and creating accountability for meeting them

2. Decentralizing responsibility for decisionmaking on instructional practice and the use of time and resources to schools and teacher teams

3. Requiring schools to implement a Comprehensive School Reform design and utilize research-based curriculum and instructional strategies

4. Reorganizing district resources to provide support and professional development revolving around each teacher's professional growth needs and each school's uniquely developed strategy.

CPS leaders – and civic leaders – see these four pieces as inextricably linked. The strong Cincinnati business community has long urged decentralization, and the Cincinnati Teachers' Union has been in the forefront of teacher decisionmaking. The business community has been equally determined that school site autonomy must be accompanied by a strong accountability system. CPS has been clear that all schools need to have a coherent, schoolwide instructional strategy that builds on research-based practice. The requirement that schools adopt a comprehensive school design allows CPS to insist on such practice and still allow a great deal of school-site flexibility in instructional decisions. Finally, from the beginning, community, union and school leaders insisted that there be a strong plan of support and intervention for the lowest-performing schools.

Intervention for low-performing schools is designed to improve instructional practice and to support districtwide reform efforts. For example, most schools that are rated Intervention schools receive Standards in Practice (SIP) training and coaches. The SIP program provides coaches to work with teacher teams to review student work and student-performance data. The SIP coaches help teachers analyze performance data to make decisions about how they should change instructional practice.

The SIP coaches are experts in content areas such as literacy and math; they also are experts in the use of student-performance data. SIP coaches receive training and support from the Education Trust's Ruth Mitchell, who developed SIP methods and refined them through work with school teams nationwide. The SIP coaches work with schools for at least one day a week over a full year, an intensive commitment wholly geared to improving the quality of teaching. The program has received high marks from evaluators and teachers for its focus on instruction.[15] The SIP program creates an important link between support and accountability that makes it possible for schools to move quickly and for the district to insist on change when necessary.

The SIP process connects the major pieces that support schools. SIP coaches work closely with the master-principal and master-teacher intervention teams that are assigned to schools placed in the Intervention category. As with the SIP coach, the intervention team spends at least one day a week working in schools for a full year. The SIP process works hand in hand with

Comprehensive School Reform (CSR) designs as well. In some cases, the consultant who provides CSR design support receives training in SIP and leads the process in schools. This ensures the complete integration of multiple approaches to school improvement.

Principle Two: *Schools are the primary units of accountability. Student achievement must be the dominant measure of school performance, but other metrics should count. Schools must be held accountable for the performance of all students (with minimal exemptions) as well as the performance of selected student groups. Changes in performance over time as well as performance levels should be measured. Also, all other functional units of the district should be held accountable for performance.*

The Cincinnati strategy requires all schools to implement a comprehensive school design from a prescreened list approved by the district, which has demonstrated effectiveness. These designs vary widely from one another in curriculum materials, schedules, staff and use of formative assessment tools, but the district judges whether each design has met its threshold for evidence of effectiveness. Regardless of design, CPS requires that all schools demonstrate their plan meets high standards for instruction and learning.

The district developed 10 standards, called Students First, for quality instructional programs and teaching. The district also developed rubrics for measuring how well schools meet the standards. The Students First standards require all schools to implement schoolwide a proven literacy approach, provide small-group and individual instruction when needed based on student performance, and use performance-based assessments to measure student progress throughout the year. Schools develop annual plans called One Plans that describe schoolwide efforts in each area of the Students First standards and include a self-assessment of how well they meet high standards of instructional excellence. [16]

Principle Three: *All students should be assessed at every grade level every year by standardized test that are aligned as closely as possible with the curriculum, cover as much of the curriculum as possible, and validly and reliably measure what children know. Also, assessments must be fair and seen to be fair.*

CPS uses student achievement as measured by the Ohio Proficiency Test as the basis of its accountability system. These tests are administered in grades 4, 6, 8-10 in reading, writing, math, history and citizenship. Off-Grade Proficiency Tests are administered in 2, 3, 5 and 7. Thus, students are tested annually in core subjects.

Principle Four: *Accountability means consequences, both positive and negative, for schools and for everyone in the system. Everything about the accountability system – structure, process, information about assessments, assessment results, accountability ratings, consequences and more – should be broadly communicated in easily understood language to all district employees, parents and the public.*

The district rating system measures a school's current performance and improvement. The ratings are based on scores on the Ohio Proficiency Tests (OPT), attendance and dropout rates. These can be justified to the public as valid indicators of school performance. Based on these measures of performance, schools are classified into one of four categories: Achievement,

Improvement, Intervention and Redesign. The rating categories, as shown in Figure 2, are descriptive and convey a course of action for a school based on its rating.

Figure 2: CPS Composite Rating Matrix

	School Incentive Award		
Improving	Achievement	Achievement	Achievement
Improving Slightly	Intervention	Improvement	Improvement
Not Improving	Redesign	Intervention	Improvement
	Below Average Status	*Average Status*	*Above Average Status*

Performance Improvement Rating (vertical axis label)

Performance Status Rating

The district rates schools based on improvement of OPT scores over time, as well as absolute achievement levels. A school's scores are plotted on the grid below to determine the final accountability rating. Reaching school targets is the most important factor in determining the school rating. Schools that reach their improvement targets receive the highest rating – Achievement. For schools that do not reach the target but improve in some areas, ratings depend on how school performance compares to the district. If the school performs at the district average or above, it is rated as Improvement, the second-highest category. If the school improves somewhat, but performance is still below the district average, then the school is rated as an Intervention school. If the school does not reach more than 33% of the target, then the school will be rated in the lowest category and will be eligible for redesign.

Over the last five years, CPS has grappled with the details of the measurements and methods used to rate schools. It is these measurement details that determine whether the ratings reflect real improvement in student performance. In creating the metrics, CPS has had to consider (a) how best to measure student performance given their statewide test, the Ohio Proficiency Test; and (b) how much student performance to target. The district uses school average scores to measure both improvement and absolute performance. The simplistic use of averages, however, does not measure individual student growth. Average scores do not account for:

1. Changes in student population from one year to the next that could occur because of high mobility, neighborhood changes or simple statistical variance in class composition
2. The need to improve instruction for all students in all race and achievement categories
3. The possibility of "gaming" the system by not testing students likely to perform poorly on tests, such as special education students.

To help make the ratings more accurate, CPS implemented a set of revisions, which will be reflected in the 2003-04 school ratings. First, the district will measure individual student gain from one year to the next for grades 3-8. This means that instead of measuring whether students meet a pre-set passing standard, the district will measure the total amount of individual student improvement. In its first year, this will be one of the indicators used to measure performance, but the district will continue to evaluate its use.

The second major adjustment involved adding an indicator to measure improvement in subgroups of students by racial category. Schools can now earn points by reducing the gap in achievement scores by race. Schools will only gain points in the rating system if the lower scoring group improves.

Finally, the district continues to evaluate ways to encourage schools to exempt fewer students from testing and to test more students in general. As it has from the beginning, CPS includes the scores of exempted special education or limited English students if the scores help the school, but not if they hurt the school ratings.

CPS also has refined its method for measuring improvement over time. School targets require 4-5% more students to pass the OPT each year. The district now uses a three-year average as the baseline so schools that have large increases in one year are not penalized. Schools earn points based on meeting the target. Schools receive partial points if achievement gains are at least 33% of the established target. All schools must improve at the same 4-5% rate, *except* those that have reached the district's "threshold" rates. For example, the district set a threshold passing score of 90% for the Ohio Proficiency Test. If schools reach this target, they automatically receive full points regardless of reaching the improvement target. Threshold targets include:

Ohio Proficiency Tests	90% Passing
Off-grade Proficiency Tests	90% Passing
Student Dropout Rates	2.5%
Student Attendance	93%

The district added thresholds to the accountability system in response to outcry from relatively high-performing schools that were unable to improve performance to the same extent as low-performing schools. These schools felt penalized by their high performance, as they could never reach the Achievement category, even though they had the highest performance in the district.

The district rating system is consistent with – but more demanding than – the state rating system. The state's district rating system is based on scores on the Ohio Proficiency Test in reading, writing, math, social studies and science along with other indicators such as graduation and attendance. The state rates districts based on the number of indicators on which they achieve target levels.

Also, principals and teachers have performance evaluations that are tied to the school accountability system. The teacher salary structure for new teachers is actually tied to teacher

proficiency, measured, in part, by student achievement. The district did create a plan that linked all compensation for all teachers to demonstrated proficiency, but the Cincinnati Federation of Teachers voted to rescind this plan in 2002.

Schools that achieve 75% or more of their potential points receive the School Incentive Award. Eleven schools received such awards in 2002 (based on 2001-02 performance). Teacher and principals and other professional staff in these schools each receive $1,400 bonuses and other staff receives $700 each. This represents a significant commitment of resources. The district budgets over $300,000 each year to pay for these bonuses. In addition, these schools also earn greater autonomy. Schools that reach this level can opt out of Off-Grade Proficiency testing. Last year, the district also gave these top schools their pro-rated portion of dollars slated for professional development at the district-run Mayerson Academy to use at their discretion. The district continues to look for other ways to give these schools increased autonomy and flexibility.

Schools in the bottom two categories also receive significant resources. The lowest-performing schools receive assistance from an Intervention team or become eligible for school redesign. CPS spent just under $1 million in the year 2000-01 to support schools in redesign or intervention. [17]

Redesign schools are reorganized to implement research-based comprehensive school designs such as Direct Instruction, America's Choice and Expeditionary Learning.[18] All existing teachers are surplused, meaning they are put into the district hiring pool, and new teachers are hired to fit the new structure and philosophy of the school design.

Intervention schools, the second-lowest category, receive intensive support from a two-person Intervention team composed of a master teacher and master principal. The team conducts an in-depth review of the school's instructional conditions and has broad authority to recommend changes in the school. With the approval of a district-level Redesign Committee that includes top union and district management, the team works closely with these schools over the year to help make improvements.

The Intervention team conducts an intensive set of evaluation visits using the Students First rubric described earlier; all schools use this rubric in comprehensive planning to guide them in assessing the instructional conditions in the school. The Intervention team then develops a set of recommended changes to be included in the school support plan for the year. These changes can include reassigning or retraining staff, changing the schedule or changing course offerings.

Future Challenges
Through the accountability system and its associated pieces, the district is beginning to achieve improved success. In some cases, as with the two Redesign schools that recently won top honors, this success has been dramatic. The move to measuring student gains should give the district a more accurate sense of real improvements. In addition, it may help more accurately reflect the improvements in some of the most troubled schools where the district's methods did not adjust enough for the student mobility. The practice of measuring student gain may also put pressure on

schools that have already reached the district's top threshold levels, thus invigorating the system for schools at the top end as well.

But, the key challenge for CPS lies ahead. As for all districts, the true test will be whether CPS can develop ways of understanding why some schools improve and others do not and then *act* to support and change schools in ways that really work. Analysis of school status over time shows that some schools have been able to improve and do so dramatically. Other schools, however, have dropped back down or stayed at low levels of student achievement over the past four years.

In Figure 3, the chart shows how school ratings for elementary schools changed in the last four years. Of the five schools that were rated as Redesign in 1999, three have moved up in categories, two to the top category. Intervention schools did not fair as well: fewer than half moved up and four schools actually declined into Redesign over four years. The challenge highlighted here is to find ways of helping schools on the edge of failure. There is also challenge at the top end. Twelve schools received CPS' highest rating in 1999, and only five of them remained there in 2001. This raises questions about how to define high performance that lasts and what kinds of resources are required to keep it there.

Figure 3: Changes in CPS Elementary School Accountability Ratings, 1999-2002
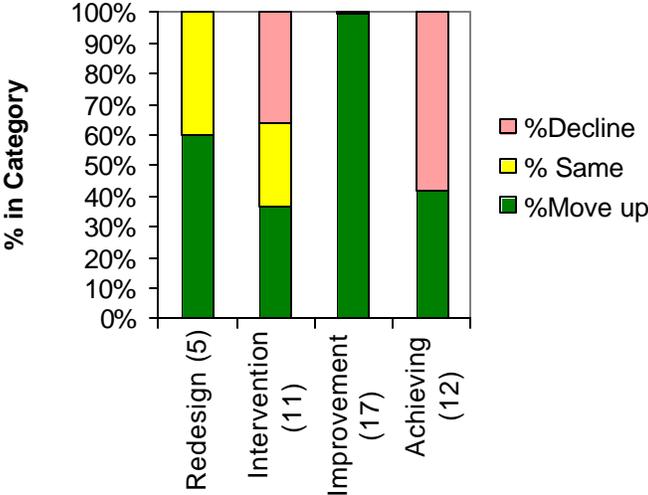
Though all involved know the system isn't perfect, there is one thing that is clear. The accountability data, combined with CPS' very public process of allocating resources to schools, have given district, community and union leaders a place to begin asking the questions that matter.

Case Study #3 – Houston Independent School District: An Improving Accountability System To Support Whole-Systems Change

Executive Summary
School accountability in Texas began with the assessment of state standards by the Texas Assessment of Knowledge and Skills (TAAS). The Houston Independent School District (HISD) established an accountability system in 1993-94 that was designed to extend the power of the

Texas accountability system. Like the Texas system, the HISD system used TAAS as its primary measurement, but unlike the Texas system, the Houston system focused on performance improvement as well as performance level.

In 1998, HISD made an important change to its system, increasing significantly the incentives for schools to test every child. As a result, in 1998-99 exemptions were significantly reduced, and for the first time, school performance decreased from the previous year.

In 2001, the school board made an even more significant change. It added to the TAAS a second measure of student achievement, the Stanford 9, a norm-referenced achievement test that includes reading, language, mathematics, spelling, science and social studies. Results are given in percentiles, with the 50th percentile being an average score. For Spanish-speaking students, Aprenda 2 was added.[19]

TAAS was weighted 70% and the Stanford/Aprenda was weighted 30 percent. In addition, high schools were given incentives to increase the percentage of seniors graduating with either the Recommended High School Program or the Distinguished Achievement Program and the percentage of juniors taking the PSAT. The new accountability measures were fully implemented in 2002.

At the same time, HISD began phasing in a new budget/management system built on weighted student funding and implementing a major, comprehensive online curriculum and student-information management system. Weighted student funding, with the money following the child, increased school empowerment, and the new curriculum and student-information management system increased capacity for excellence. These systems were designed to support HISD's strategy for whole systems change, which rests on reciprocal accountability, matching the pressure of the change in expectations with increased capacity and empowerment.

HISD's progress in improving student achievement was recognized in July 2002 with the award of the inaugural Broad Prize for Urban Education.

Narrative
In 1989, four new trustees were elected to the nine-person HISD Board of Education. Soon after these board members took the oath of office in January 1990, a fifth new trustee was appointed by the board to fill a vacancy created by an unexpected resignation. The five new board members, working with some returning board members, soon determined to launch HISD on a bold reform path.

In June 1990, the board adopted a vision statement, *A Declaration of Beliefs and Visions*, that mandated reform under four headings: (1) HISD exists to support the relationship between the teacher and the student; (2) HISD must decentralize; (3) HISD must focus on performance, not compliance; and (4) HISD must require a common core of academic subjects for all students.

Explicitly stated under the second and third headings were a commitment to placing decisionmaking as close as possible to the teacher and the student; establishing clear goals, high standards and effective systems of evaluation; holding schools accountable for results; evaluating

schools based on improvement trends; and providing specific incentives to reward improvement. Without fully understanding the implications, the HISD board had committed HISD to a school district accountability system within the framework of whole systems change.

Much happened during the next three years: the superintendent was replaced; shared decisionmaking committees were established at every school; administrator evaluations became more realistic; principals were given professional assistance so they could more easily manage (and if necessary terminate) inadequate teachers; and the board of education angered the business leaders and almost everyone else by approving, in the heat of summer 1992, a 32% increase in local property tax rates. Nothing was done, however, to make schools accountable for student achievement.

This was partially because HISD had no way to assess student achievement. The state kept changing the state assessment – the Texas Assessment of Academic Skills (TAAS) – making it impossible to lay down a baseline and measure changes in performance. In anticipation of the state assessment, HISD had dropped the use of its previous measure, the Metropolitan Achievement Test. Also, perhaps because he anticipated resistance from the workforce, the superintendent, who had been hired in 1991 to implement *Beliefs and Visions*, seemed in no hurry to act.

The business community and the board, however, were in a hurry. Business leaders had demanded, during the tax battle of 1992, that accountability precede a big tax increase. And board members, impatient with two years of delay and feeling the heat from the business community, made the establishment of a district accountability system their priority for 1993. Still, by late summer 1993, the superintendent had nothing to recommend.

Meanwhile, with TAAS sufficiently in place to support accountability, the 73rd Texas Legislature, in spring 1993, had established the framework for a state accountability system. The accountability measures were complex. School districts and schools were to be measured by student TAAS performance levels, dropout rates and attendance rates. Performance indicators were to be disaggregated by race or ethnicity, sex and socioeconomic status. District and school performance and performance ratings would be published in an annual report, and campus report cards would be provided to parents. Districts and schools that failed to meet minimal standards were subject to a list of sanctions.

In late August 1993, the HISD board demanded action on a district system. With very active board leadership, staff committees began to set forth principles, collect data and draft documents. Though it took six months to fully develop the HISD accountability system, the major policy decisions were approved by the board on October 7.

The HISD accountability system changed somewhat from 1993 to 2001, as did the Texas system. Both gradually raised the bar for expected performance, but both remained quite stable. The HISD system reflected the Texas system. It used the same performance indicators and the same weights for performance indicators (100% TAAS for elementary schools, and for middle and high schools, 90% TAAS and 10% attendance and dropout rates) to rate schools based on

performance level as *exemplary, recognized, acceptable* and *low performing*. Unlike the Texas system, HISD divided *acceptable* into two performance levels, *acceptable* and *low acceptable*. (The HISD board wanted a higher standard for *acceptable*.) Also, unlike Texas, HISD did not disaggregate performance indicators by student groups. Given the demographics of HISD – 90% children of color and 57% (rising to nearly 80%) receiving free or reduced-priced lunch – it seemed an unnecessary addition of complexity.

The main difference between the HISD and Texas systems was that in addition to performance level, HISD measured performance change. For schools in each of the five performance levels, HISD established expected gains in TAAS performance. Lower-performing schools were expected to make greater gains than higher-performing schools. The expected gains were: 8% increase for schools with pass rates 20% and below; 6% for schools with pass rates between 21 and 39%; 4% for schools between 40 and 64; and 2% for schools between 65 and 89. Schools that achieved the standard for exemplary, 90%, were expected to stay above 90%.

Schools that achieved their expected gain received an *acceptable progress* rating. Schools that doubled their expected gain received a *recognized progress* rating. Schools that tripled their expected gain received an *exemplary progress* rating. There were also two ratings – *unacceptable progress* (later *minimal progress*) and *no progress* – for schools that made less than expected gains or slipped backward.

By bringing together the five ratings of performance level with the five ratings of performance progress into a matrix, with performance level on the horizontal axis and performance progress on the vertical axis, HISD was able to show the performance level and performance improvement of every school in the district and the district as a whole. Every school was in one of the 25 cells of the matrix. One could identify quickly the schools in the bottom right of the matrix, low-performing schools making unacceptable progress or no progress at all, and schools in the top left of the matrix, schools improving rapidly and thereby moving into higher-performance levels. Also, by how heavily the matrix weighted toward the bottom right or top left, one could see the overall performance of the district. The district could see immediately which schools deserved recognition and rewards for performance level or performance improvement and which schools needed immediate intervention.

Figure 4: HISD Accountability Matrix

	Current Progress Rating				
Current Performance Rating	**Exemplary**	**Recognized**	**Acceptable**	**Minimal**	**No Progress**
Exemplary >90%					
Recognized 75-89%	6% improvement	4% improvement	2% improvement	improvement less than 2%	No improvement
Acceptable 60-74%	12% improvement	8% improvement	4% improvement	improvement less than 4%	No improvement

Low Acceptable 50-59%	18% improvement	12% improvement	6% improvement	improvement less than 6%	No improvement
Low Performing <50%	24% improvement	16% improvement	8% improvement	improvement less than 8%	No improvement

The first matrix, in October 1993, which showed performance for the 1992-93 school year, was a revelation. Eight elementary schools had TAAS pass rates of 90% or more. Four were located in upper-middle-class residential communities, but had significant minority enrollments. The other four schools were in poor neighborhoods and were predominately minority and poor. Twenty-nine elementary schools had increased their TAAS pass rate by 12% or more. Almost every one of these was in a poor neighborhood.

In January 1994, the board chose board member Rod Paige to be superintendent. Over the next five to six months, the Paige administration developed a comprehensive system of rewards and recognition for exemplary and recognized schools and interventions, responses, and sanctions for low-performing schools. Rewards were mostly increased freedom at the school site and recognition mostly the spotlight and special events. Interventions were mostly intervention teams, additional resources and changes in personnel. Sometimes a significant number of teachers joined the management team in being reassigned. No schools were formally reconstituted.

Though standards for performance were raised each year, and numerous small changes were made to the system of recognition, rewards and interventions, the HISD accountability system remained basically unchanged until the 1998-99 school year. In spring 1998, the board, responding to a loophole in the system, voted to tighten the standard for exemptions.

Not anticipating attempts to game the system, the board had accepted the state definition of who should be tested. For accountability ratings, the state used the pass rate for non-Special Education students tested on the English TAAS enrolled as of the state-established fall enrollment date. Principals all over the state were exempting large numbers of students from the TAAS, mostly students enrolled in bilingual education. Some HISD schools with recognized status had exempted 60 or 70% of their students from the TAAS. Some school districts in Texas were making a mockery of the state accountability system.

The HISD board solved this problem with a simple policy change. Henceforth, with a very limited exemption for children who could not speak English and had been in the United States less than a year (many of whom had not been educated in any language) and severely disabled students who could not be tested, every child counted. As Superintendent Paige put it, "Every child who can hold a pencil will be tested." To establish a school's accountability rating, instead of counting the number of children who passed the TAAS as a percentage of those who took the test, HISD counted the number of children who passed either the English or Spanish TAAS or the SDAA (the off-grade-level test for Special Education students) as a percentage of the children enrolled, less the small number of children "who could not hold a pencil." Schools now had an incentive to test every child possible.

In response to this policy change, school performance in 1998-99 dropped for the first time since the establishment of the system. Since 1994, which was the best baseline year – 1993 was the first year the TAAS was given at all grade levels – the shift from low performance to high performance had been impressive. The number of *exemplary* schools had increased from 10 to 84 and the number of *low acceptable* and *low performing* schools had dropped from 78 to 2. But in 1999 the number of *exemplary* schools dropped to 10, the number of *low acceptable* schools increased to 48, and the number of *low performing* schools increased to 5.

In February 2001, the board approved an even more significant change in the HISD accountability system. It added student achievement on the norm-referenced Stanford 9 or Aprenda 2 as a second measure of school performance, and, for high school performance, added both the percentage of high school seniors taking either the recommended or advanced Texas curriculum and the percentage of juniors taking the PSAT. These changes were developed by a Committee of the Whole of the Board of Education, supported by district administrators and researchers, including school principals. Numbers were produced in summer 2001 to show schools how they would perform under the new system, but the new system did not take effect until 2002.

The impetus for these changes came from the board, which wanted multiple measures of student achievement and wanted high schools to push students into more rigorous courses. In anticipation of adding a second measure of student achievement, the board had prepared the way for using the Stanford 9 and Aprenda 2 for accountability by asking the administration in 1996 to begin administering these two tests every year at every grade level.

The new measures require some explanation. For elementary schools, the only change was that the TAAS, used in exactly the same way it had been used before to determine performance level, was weighted 70% of student achievement. The Stanford 9 and Aprenda 2 (for Spanish-speaking children) were brought together into a second measure of performance and weighted 30%. For the Stanford/Aprenda measure the standard was the number of children scoring at or above the 50th national percentile rank on the Stanford 9 Complete Battery or the Aprenda 2 Basic Battery. *Exemplary* was 70% or above of the children meeting this standard; *Recognized,* 60%; *Acceptable,* 50%; *Low acceptable,* 40%; and *Low performing,* below 40%.

The expected gains on the Stanford/Aprenda for the progress rating were: 8% for schools with fewer than 40% of their students meeting the 50th or greater percentile standard, 6% for schools with between 40 and 50% meeting the standard, 4% for schools with between 50 and 60% of their students meeting the standard, and 2% for schools with more than 70% of their students meeting the standard. Schools with 70% or more of their students meeting the 50th percentile standard were expected to maintain performance above this level.

As with the TAAS, schools could achieve a *Recognized* or *Exemplary* progress ratings by doubling or tripling the expected gain, or they could fail to meet the expected gain or show no gain and be rated *Minimal* or *No Progress*. For performance level, the TAAS is weighted 70% of student achievement and the Stanford/Aprenda 30%. For performance improvement, the TAAS

and Stanford/Aprenda were weighted equally. For an HISD Research Brief describing the 2002 Accountability System in detail, see Appendix I.

In addition to these changes in measuring student achievement, in 2001 the board added a second Scholars Award for high schools. High schools were designated as *Platinum, Gold, Silver or Bronze* based on the percent of graduates who were designated Texas Scholars by completing either the Recommended High School Program or the Distinguished Achievement Program and the percent of juniors taking the PSAT. Four points were awarded to high schools with more than 75% of graduates designated Texas Scholars, 3 points for more than 50%, 2 points for more than 25%, and 1 point for less than 25%. Four points were awarded to high schools with more than 75% of juniors taking the PSAT, 3 points for more than 50%, 2 points for more than 25%, and 1 point for less than 25%. Schools with 8 points received the *Platinum* rating; schools with 6-7, *Gold*; schools with 4-5, *Silver*; and schools with 2-3, *Bronze*.

The board established the Scholars Award in response to data showing that more than 60% of all HISD graduates were graduating with only the minimum-required state high school curriculum, and very few high school juniors were taking the PSAT. In a few years, the board indicated it might wish to count PSAT scores as well as PSAT participation in the Scholars Award. The board decided to place these measures of performance outside the regular accountability system rather than weight them into the system and dilute other measures of performance. In this way, high schools have the opportunity to be *Exemplary Platinum*, *Exemplary Gold* or any other combination of recognition. The incentive to perform on all measures remains high.

The new accountability system, as expected, depressed school performance. Table 7 shows changes in school performance ratings from 1994-2002.

Table 4: HISD Accountability Performance Ratings, 1994-2002

Ratings	1994	1995	1996	1997	1998	1999	2000	2001	2002
Exemplary	10	19	43	53	84	10	20	36	24
Recognized	56	68	100	112	117	63	107	107	62
Acceptable	86	82	68	65	50	125	106	109	156
Low Acceptable	70	67	31	13	0	48	16	0	23
Low Performing	8	2	0	0	2	5	2	0	4
Total	230	238	242	243	253	251	251	252	269

The revised HISD accountability system was developed and phased in concomitant with two other major reform initiatives: weighted student funding and a comprehensive districtwide curriculum and student-information management system. Also, in the late 1999s, the district began to develop performance dials for all major functional units of the district and phased in major increases in individual cash rewards for achievement. Both Paige and Kaye Stripling, who replaced Paige as superintendent in 2001 understood and were committed to whole systems change. Both, along with the board, recognized that accountability must be matched with empowerment and capacity for excellence.

The weighted student funding budget and management system was approved by the board in 2000, following a year of study and extensive policy recommendations from a blue-ribbon committee of community leaders and district administrators. It called for a three-year phase-in of a system that would (1) increase the percentage of HISD dollars spent at schools to 82%, (2) replace budgeting for programs and staff with budgeting based on weighted student dollars, and (3) give principals significantly more control over how school budgets are spent. HISD principals, who already had considerable control over budgets and personnel, were now to be given the freedom, almost, of a charter school. Principals would be free to configure the workforce of their schools to meet the needs of their students, set schedules, purchase goods and services from within or without the district (within guidelines), and carryover gains and losses.

This new budget/management system required major changes in central office operations, new internal auditing systems, training for principals and even the establishment of a training program for school business managers. Also, because the new budgeting system exposed significant inequities in school funding, the new system required significant transfers of money from schools that had been overfunded in the past (though no school considered itself overfunded) to schools that had been underfunded. For all of these reasons, administration proposed, and the board accepted, a three-year phase-in.

The first phase-in year was 2000-01. But after one year, principals and central office administrators determined that the transition was confusing. Superintendent Stripling recommended a budget for 2001-02 that completed the phase-in. Much administrative work remains to be done, and schools are still charged for average teacher salaries instead of actual teacher salaries, but weighted student funding is fully in place in HISD.

Weighted student funding increased equity and further empowered schools. PASS and Project Clear built additional capacity for instructional excellence. PASS for Profiler for Academic Success of Students is a comprehensive student-information management system. It enables teachers and administrators to obtain on demand data on student achievement sliced and diced any way needed.

Project Clear is a comprehensive district curriculum. Its major components are:

- Syllabus Planner: Interactive materials provide details on the scope, sequence and pacing of units for the entire curriculum. Additional instructional resources, tools and concepts also are available.

- Model Lessons in Core Subjects: Closely aligned with Texas standards, model lessons include assessment strategies and curriculum activities. Model lessons for 13 core courses are being developed in 2002-03. An additional 14 content areas/courses will be available in 2003-04.

- Professional Development: Lead teachers were trained during mid July and early August of 2002. They are now training their peers in the schools. In addition, multimedia online-linked training modules are becoming available. Teachers can use district provided laptops to log on to the training system during off hours.

- Curriculum-based Snapshots of student assessments for diagnostic purposes: Through the HISD Connect Web Portal, teachers can download and administer snapshots to measure student mastery of the syllabus materials.

Performance dials sprang from an operations improvement strategy launched by Paige in his first year as superintendent. Peer Examination, Evaluation and Review (PEER) was a comprehensive systems improvement strategy built around task forces made up of HISD employees and experts from the business community. These PEER teams applied standard quality improvement tools to assess customer satisfaction and process control for most major functional areas of the district. Nearly 20 PEER teams and the implementation of hundreds of recommendations significantly improved district operations.

By the late 1990s, every operating unit of the district had developed performance dials that recorded measures of performance. Typically, the metrics included measures of productivity, quality, timeliness, customer satisfaction or safety, and showed performance against expectations. The performance of functional units, as shown by performance dials, became a significant part of personnel evaluations.

Paige launched the Campus Improvement Incentive Program in 1997. This program provided cash awards to campus-level employees. Schools who were rated exemplary or recognized on the Texas Education Agency (TEA) Accountability System or demonstrated exemplary growth on the HISD Accountability Systems received a campus improvement incentive.[20] All employees who worked on campuses and had met the baseline eligibility requirements shared in the award.

Following the success of the campus incentive plan, HISD implemented an aggressive performance incentive program for principals and area superintendents in 2001. This incentive pay was tied to accountability ratings[21] and outstanding progress toward meeting student performance objectives. Area superintendents could earn awards of up to $20,000. Principals could earn awards up to $5,000.

The new accountability system, the weighted student funding budget/management system, Project Clear and PASS, performance dials for operating units, and the cash incentive system were designed to align with each other in mutual support of the district's whole systems change strategy. Together they mark the most recent stage in HISD's implementation of the 1990 *Declaration of Beliefs and Visions.*

What is the result of this decade of accountability-led whole systems change? In July 2002, the Broad Foundation announced that HISD was the winner of the inaugural Broad Prize for Urban Education. Among the reasons given by the jury that selected HISD for this prestigious award were the following:

Dramatically increasing student achievement.
- Significant increases in student achievement were made in both reading and math at the elementary, middle and high school levels over the past three years.

- Gains in student achievement were made at a faster rate than expected for districts in Texas with similar poverty levels.
- Students performance at higher levels than expected for districts in Texas with similar poverty levels.

Successfully reducing achievement gaps.
- Demonstrable decreases in achievement gaps across ethnic groups were greater than any other finalist.
- Demonstrable decreases in achievement gaps between high and low-income students were greater than any other finalist.

Results confirmed by expert evaluators.
- Resources, support and professional development provide school and district leaders the necessary tools to achieve academic objectives.
- Academic objectives are clearly defined.
- School and student performance is regularly monitored.
- Rewards, intervention and adjustments to support schools are based on student performance.[22]

Observations for Policymakers

It is hoped this survey of school district accountability systems, the four principles proposed for district accountability, the work of the nine Threshold Districts and the case studies of Boston, Cincinnati and Houston will motivate district policymakers throughout urban America to establish or improve school district accountability systems. For those so disposed, some closing observations are offered.

First, develop an accountability system that manifests the core principles of district accountability and includes as many as possible of the points embedded in these principles. The more closely an accountability system reflects these principles, the greater will be its power to drive improvements in student achievement. Here again are the four principles proposed in this report.

1. *District accountability systems must be embedded in a comprehensive theory of action that includes all the elements of standards-based reform, school empowerment, and district and school capacity for high performance. And because accountability systems exist to improve teaching and learning, every part of the system must be designed with this purpose in mind and aligned to make it possible.*

2. *Schools are the primary units of accountability. Student achievement must be the dominant measure of school performance, but other metrics count. Schools must be held accountable for the performance of all students (with minimal exemptions) as well as the performance of selected student groups. Changes in performance over time as well as performance levels should be measured. Also, all other functional units of the district should be held accountable for performance.*

3. *All students should be assessed at every grade level every year by standardized tests that are aligned as closely as possible with the curriculum, cover as much of the curriculum as*

possible, and validly and reliably measure what children know. Also, assessments must be fair and seen to be fair.

4. *Accountability means consequences, both positive and negative, for schools and for everyone in the system. Everything about the accountability system – structure, process, information about assessments, assessment results, accountability ratings, consequences and more – should be broadly communicated in easily understood language to all district employees, parents and the public.*

Second, align your district accountability system with your state accountability system as much as possible. Use of common tests, terms and structures will reduce confusion. If your state system is weak, however, don't be reluctant to strike out in a different direction. As states respond to the requirements of No Child Left Behind, this issue should be less problematic, for No Child Left Behind reflects the principles proposed, and a district system built on these principles should be in perfect alignment with No Child Left Behind.

Third, don't try to do everything at once. As shown in the case studies, good accountability systems evolve over time. Start with the basics and expand and refine over time. As impatient as you might be to go fast, implement change in careful stages.

If an additional assessment is needed, first use it for diagnostic purposes for a year or two. Then begin using it for accountability. Start consequences with school accountability, especially using the spotlight. Add positive consequences before you implement negative consequences – though moving quickly to intervene in low-performing schools is a priority that cannot be delayed – and for individual consequences, start with school principals and then add over time others who are responsible for student achievement. Don't delay too long holding all functional units accountable, including individual accountability, for doing so is not fair to school people.

Recognize that as more accountability is piled on the system and as individual consequences increase, principals and teachers must be given more control over budgets, personnel, schedules and other school operations. Strong accountability will almost inevitably drive a district to give schools many of the freedoms of a charter school.

Control of the principalship, curriculum and professional development, however, cannot be turned over to schools. The district must manage its core business. Empowered schools means empowered principals working with teacher teams. It means that shared-decision committees made up of teachers and parents have only advisory responsibilities. Principals are in charge.

In short, look on accountability as a principle and a goal, and develop accountability systems over time in ever-improving iterations that are always fully aligned with all other district operations. Remember that accountability is only one link in the chain of standards-based reform and that the standards and accountability chain most be kept in close alignment with empowerment and capacity for excellence.

Fourth, know that change is frequently difficult. Expect resistance and anticipate it by doing every thing you can to build support for accountability among principals, teachers, other district

employees, parents and the public. Don't assume that people understand. Develop an effective internal and external communications plan, and keep working it and improving it. This will help minimize resistance, transform the district culture from compliance to performance, and guarantee the continuation of accountability as a guiding district principle from one superintendent and board to the next.

Finally, recognize that establishing a district accountability system is a political act. Just as state accountability systems spring from the state political process and No Child Left Behind is a federal statute, so district accountability is a reflection of the political will of a community. Principles are one thing. Politics are sometimes something else.

Consider the accountability systems included in this survey. Few are robust. Does this mean that policymakers in the other districts wanted nothing more? It is likely that in some cases policymakers wanted more and had to settle for less. The lack of knowledge by survey respondents in some districts about plans for improvement does not necessarily mean the superintendent and board members in the district are content with the status quo. They may be planning significant improvements to the district accountability system as soon as key individuals or interest groups can be won over or outmaneuvered.

Policymakers who hope to implement or improve district accountability systems need to be master politicians, as well as students of district accountability and whole systems change. This report will not help them become better politicians. But perhaps, it will help them become more informed students of district accountability and motivate them to place district accountability at the center of their strategic plans for improving student achievement in the nation's urban schools.

ENDNOTES

[1] Grissmer, David W.; Flanagan, Ann; Kawata, Jennifer; Williamson, Stephanie (2000). *Improving student achievement: What state NAEP test scores tell us.* Santa Monica: RAND. MR-924-EDU, accessed October 29, 2002.

[2] National Center for Education Statistics, (2001). *Characteristics of the 100 largest public elementary and secondary school districts in the United States: 1999-2000.* Washington, DC: U.S. Department of Education.

[3]

[4] Office of Elementary and Secondary Education (2002). *No Child Left Behind: A desktop reference.* Washington, DC: U.S. Department of Education.

[5] Philadelphia is no longer using the accountability model described in this report. Details on Philadelphia's intervention/reward/sanction structure were not available to include in this database.

[6] The state of California's Academic Performance Index (API) is a rating assigned to schools calculated from scores on the California Standards Tests and the SAT-9.

[7] Hawley-Miles, Karen (2002). *A theory of action, a work in progress: Boston Public Schools' accountability system.* Unpublished article.

[8] Menino, Thomas (1996). State of the City address.

[9] Policy Studies Associates (2002, October). *Student achievement in schools assessed as demonstrating effective practices in the Boston Public Schools.*

[10] Boston Plan for Excellence is Boston's public education fund.

[11] Boston Public Schools (2002). *Theory guiding the essentials of whole school improvement.* Whole School Improvement Rubrics.

[12] Hawley-Miles, Karen (2002). *Cincinnati Public Schools accountability system: Strategic coherence and serious consequences.* Unpublished article.

[13] See Ohio state accountability Web site at http://www.ode.state.oh.us/.

[14] Schools earning lowest rating in 1997-98 or 1998-99.

[15] Supovitz, Jonathan A. and Watson, Susan (2000). *Team based schooling in Cincinnati: The second year.* Cincinnati: The Knowledgeworks Foundation (formerly Conlan Foundation).

[16] If a school is using a Comprehensive School Reform Design that has its own planning tool, schools may substitute this for portions of the One Plan.

[17] Hawley-Miles, Karen, and Hornbeck, Matthew (2001). *Getting to the heart of improving instruction: Professional development spending in the Cincinnati Public Schools.* Education Resource Strategies.

[18] These were the designs chosen in the first four redesign schools.

[19] Aprenda is a norm-referenced achievement test specifically designed and created in Spanish, not translated, to meet the needs of Spanish-speaking students. The achievement test battery is published by Harcourt Brace and includes testing in reading, mathematics, language and listening. Results are presented in national grade equivalents, stanines and percentiles.

[20] Houston Independent School Compensation Manual 2002-03.

[21] HISD schools that were rated exemplary or recognized on the Texas Education Agency (TEA) Accountability System or demonstrated exemplary growth on the HISD Accountability System. HISD Accountability System weights the results of the Stanford 9/Aprenda and TAAS tests.

[22] Broad Foundation (2002). News release. http://www.broadfoundation.org./med-news/2002-1002-hous.shtml.